E. A. Yakubaitis

fundamentals of engineering cybernetics

TRANSLATED FROM RUSSIAN

Published for the National Aeronautics and Space Administration
and the National Science Foundation, Washington, D.C.
by the Israel Program for Scientific Translations

INSTITUT ELEKTRONIKI I VYCHISLITEL'NOI TEKHNIKI AKADEMII NA.
RIZHSKII POLITEKHNICHESKII INSTITUT SOVETA NARODNOGO KHOZYAISTVA LATVIISKOI SSR

Institue of Electronics and Computer Engineering of the Academy of Sciences of the Latvian S.S.R.
Riga Polytechnic Institute of the Council for National Economy of the Latvian S.S.R.

E.A. YAKUBAITIS

FUNDAMENTALS OF ENGINEERING CYBERNETICS

(Osnovy tekhnicheskoi kibernetiki)

Izdatel'stvo Akademii Nauk Latviiskoi SSR
Riga 1962

Translated from Russian

Israel Program for Scientific Translations
Jerusalem 1965

NASA TT F-290
TT 65-50116

Published Pursuant to an Agreement with
THE NATIONAL AERONAUTICS AND SPACE ADMINISTRATION, U.S.A.
and
THE NATIONAL SCIENCE FOUNDATION, WASHINGTON, D.C.

IPST Cat. No. 1401

Translated by Z. Lerman
Edited by S. Alexander

Price: $6.00

Printed in Jerusalem at the Sivan Press

Available from the
U.S. DEPARTMENT OF COMMERCE
Clearinghouse for Federal Scientific and Technical Information
Springfield, Va. 22151

X/18

TABLE OF CONTENTS

PREFACE

This book deals with *engineering cybernetics*, a rapidly developing new science with a wide scope of application.

The book is intended for engineers concerned with the automation of industry, transport, and agriculture. It can also be used as a textbook by senior undergraduates specializing in automation, electronics, power engineering, heat engineering, etc.

The author wishes to express his gratitude to the Candidates of Technical Sciences A. N. Sklyarevich, M. P. Vaivars, A. K. Baums, and all other co-workers who went to great trouble over this manuscript and offered valuable advice.

INTRODUCTION

1. BASIC DEFINITIONS

The development of mathematics, and especially information theory, the theory of probability, and mathematical logic, have opened new perspectives in the control of various processes in living organisms and inanimate objects.

In addition, the achievements of electronics and computer engineering in recent years have provided the technical means for solving the problems of control on an entirely new level, hitherto inaccessible. In short, this is the background for the development of a new science which Norbert Wiener called in 1947 *cybernetics*.

The term "cybernetics"* is not new in science. Plato used the word to describe the art of steersmanship and, later, the French physicist Ampère used it for the art of government.

N. Wiener imparted a new meaning to this term, defining cybernetics as "*the field of control and communication theory whether in machine or animal*".

Cybernetics is a new branch of science embracing the achievements of various older provinces: physiology of higher nervous activity, theory of automatic control, mathematics, etc. It has opened new vistas for the solution of very important problems.

The various concepts constituting the basis of cybernetics were largely developed by C. Shannon, I.P. Pavlov, A.N. Kholmogorov, W. Ashby, and other scientists from all over the world.

In the Soviet Union, among the prominent cyberneticists, we have A.I. Berg, S.A. Lebedev, A.A. Dorodnitsyn, A.A. Fel'dbaum, V.V. Solodovnikov, A.G. Ivakhnenko, and V.M. Glushkov.

The efforts of A.I. Berg and his followers have made cybernetics a principal branch of science in the Soviet Union, opening wide horizons for the development of the national economy.

Engineering cybernetics is a division of cybernetics studying the design of machines capable of performing composite logical functions of automatic control. These machines, or more precisely systems, are now called *cybernetic*.

The pre-cybernetic period was mainly characterized by the automation of *manual* labor. Engineering cybernetics now enables us to proceed directly with the automation of the *mental* activities of man. This automation can be applied only to a problem for which the laws of operation can be ridigly stated, or, in other words, for which a solution in the form of an *algorithm* is available. The range of problems for which algorithms are available is growing daily.

A characteristic feature of cybernetics is the drawing of *analogies* between the activities of living organisms and the operation of a machine.

*Κυβερνητικη in Greek means "steersmanship".

These analogies, in spite of the profound difference between the nature of the living organism and that of a machine, lead to the imitation of various human functions in cybernetic systems.

The main factor involved in the design of cybernetic systems is the receiving, processing and transmitting of information.

By information is meant the multiplicity of data which man receives through his senses from his environment. For a cybernetic system information comprises messages concerning the variation of characteristics of the controlled member (a mill, a machine, a mechanism) and external conditions affecting this member. The analysis of this information, and methods of its measurement and processing, led to the development of the information theory.

A large class of cybernetic systems can be designed with the faculty of adaptation or self-adjustment. This faculty makes possible the automatic adjustment of an object whose properties vary with time.

Cybernetic systems, like living organisms, always operate with feedback. Feedback here, as in ordinary automatic-control systems, ensures exceptional flexibility in response to the variation of external conditions affecting the cybernetic system.

In the great majority of cybernetic systems the behavior is random owing to the [random] variation of external conditions. Engineering cybernetics therefore freely draws on statistical theory (theory of probability), to allow for the effect of these random changes.

Relationships between individual parameters are as a rule nonlinear, and require nonlinear-system analysis.

In spite of the complexity of cybernetic systems, they are usually analyzed by the same methods applied to noncybernetic systems. In cybernetic systems these methods are naturally developed to a higher degree.

2. TRENDS IN ENGINEERING CYBERNETICS

Three principal trends of development are observed in engineering cybernetics.

a) Engineering cybernetics in the compound automation of industry

A strong factor aiding in the application of cybernetic systems to the compound automation of industry is the availability of high-speed computers with large memory storages and equipment capable of performing logical functions.

These systems, even at the present level of development, are adequate for setting up fully automated assembly lines, shops, and plants. They can produce a product without human intervention and regulate its production so as to obtain the highest possible quality at the lowest cost.

The electronic dispatcher of the power grid is an example of a cybernetic system. This dispatcher regulates the output of electric power from various power stations connected into a single grid. It allows for the efficiencies of steam boilers, turbogenerators, the cost of fuel burnt at

thermal power stations, the water reserves at hydroelectric stations, the cost of transmission of electric power from the power station to the consumer, etc., so that the cost of power is minimum.

b) Imitations of living organisms

The study of the activities of living organisms and their imitation in cybernetic systems have led to the development of synthetic "animals" which at present are rather like toys. However, even now the processes of machine "learning" are being investigated to help in devising complex "learning" cybernetic systems.

In 1929 an electronic "dog" created by the Frenchman Henry Pirot was exhibited at the Paris Radio Exhibition. This "dog" could follow a flashlight. It "barked" and "turned aside" when the flashlight was placed too near it. In 1939 another electronic "dog" was scheduled to be exhibited at the New York World's Fair. However, on the eve of the Fair the dog was attracted by the headlights of a car and run over.

Other, more complicated, types of synthetic "animals" have been created in recent years.

Walter's "turtle" performed a set of motions which imitated some basic behavioristic features of animals. When the battery feeding the "turtle" was well charged, it acted "satisfied" and escaped from light. When the battery was low, the "turtle" looked for the "feed trough" — a charging device illuminated from above. The "turtle" could also avoid obstacles placed in its path.

The last electronic "animal" created by Walter went through a process similar to the development of a conditional reflex in an animal.

Other electronic "animals" include Shannon's "mouse", and Ducroque's "foxes", and research in this direction opens ever new and interesting possiblities in imitating the functions of living organisms by cybernetic systems.

c) Synthetic "organs" for human beings

In his play "RUR", the outstanding Czech writer Karel Čapek described mechanical "people", calling them "robots". Since then, the term robot has been applied to all mechanisms of anthropoidal external appearance which imitate human functions.

In the 18th century Albertus Magnus built an iron robot which could open a door and bow to all who entered. Many other robots have since been built. The design of robots has so fascinated the West that the creation of a universal cybernetic system, a "perfect robot", which will perform all the functions of man, is often contemplated.

Such an ideal robot is scientifically unsound, since the functions of living organisms can never be completely duplicated in a machine.

However, cybernetic systems can successfully reproduce individual functions of various human organisms and the efforts of scientists are now being concentrated on their development.

A. Kobrinskii, with a group of scientists, has created an artificial arm controlled by the muscular biocurrents of a man; artificial eyes, artificial hearts, and artificial kidneys are all among the more immediate cybernetic biological projects.

These first steps taken already indicate the future promise of this field.

3. FEEDBACK

Automatic-control systems are often called feedback systems, because in any automatic-control system* the direct coupling (Figure 0.1) between the input (the control signal) and the output (the controlled variable) is supplemented by feedback between output and input. Feedback allows for change in the controlled variable (parameter) and introduces corresponding corrections into the control signal at the input.

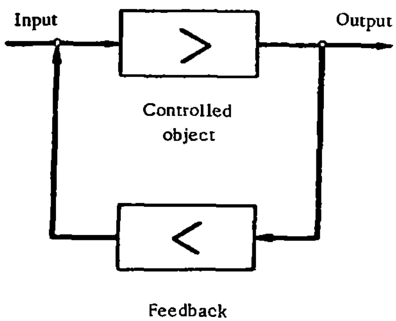

FIGURE 0.1. Feedback in an automatic-control system.

The importance attached to feedback becomes obvious when we remember that it occurs not only in automatic-control systems but also in living organisms, for both operate (or function) under variable conditions and must react appropriately to these changes.

Let us consider, for example, the function of the motor muscles in the leg of a man trying to keep his balance.

To maintain balance, the tension of various skeletal muscles must be regulated. This regulation is effected by the spiral cord (Figure 0.2) which issues appropriate "commands" to the muscles. The information concerning the balance of the body (i.e., its deviation from the normal position) is transmitted to the cerebellum. The cerebellum processes this information and gives the necessary "instructions" to the spinal cord. Following these "instructions" the spinal cord modifies the commands issued to the muscles. The spiral cord, the muscles, and the cerebellum thus constitute a closed-loop system which controls the man's balance. In this system the cerebellum is the element which closes the feedback loop between output and input.

* An exception to this rule is provided by programmed systems; these systems, however, are not cybernetic, and they operate according to a pre-ordained fixed program.

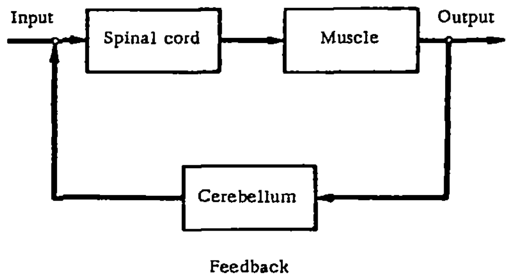

FIGURE 0.2. Block diagram of the system regulating the balance of a standing man.

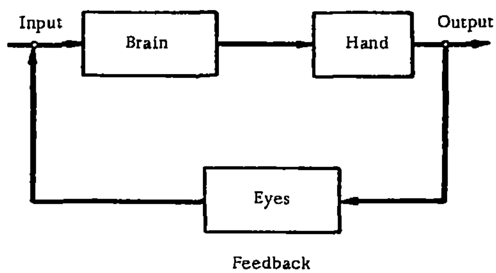

FIGURE 0.3. The human eyes as a feedback element.

Consider another example of feedback. Suppose that a man wishes to press a push button. The brain (Figure 0.3) issues the necessary signals to the arm muscles and the arm moves toward the button. The eyes follow the motion of the arm and report to the brain, which corrects the functioning of the arm muscles. The hand thus touches the push button accurately. If the eyes are closed (the feedback loop is "broken") the hand will not reach the button. It is, of course, possible to find the button by "feeling the way". But this too involves feedback by the tactile organs.

The function of the feedback loop in automatic-control systems is similar. Consider for an example the control of an electric motor (Figure 0.4). In this system the logical computing element measures the parameters of the motor shaft (rpm, acceleration, torque), makes the necessary computations, logical comparisons, and ensures control of the magnetic amplifier, which, in turn, adjusts the motor parameters.

In both living organisms and automatic-control systems, the elements have, as a rule, a certain **directionality**. The eyes transmit information to the brain (Figure 0.3), and there is no reaction from the brain to the eyes (in the same channels). Similarly, in Figure 0.4 the magnetic amplifier controls the electric motor, but the motor has no effect on the amplifier parameters. This directionality applies to any feedback loop, which has unidirectional coupling from output to input, but does not have input—output feedback.

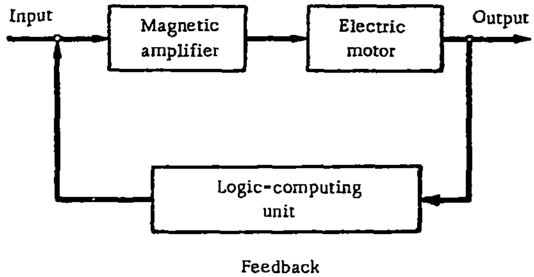

FIGURE 0.4. Feedback in a motor control system.

4. PRINCIPAL PROBLEMS OF THE THEORY OF CYBERNETIC AUTOMATIC-CONTROL SYSTEMS

The problems dealt with by the theory of cybernetic automatic-control systems are divided into two distinct categories:

a) Analysis of control systems

In this case the control system is given and it is required to determine the behavior of the system (stability, response time, etc.) for any input signal and the variation in the parameters of the controlled member and of other elements of the system.

b) Design of control systems

This problem is more complicated, since it is required to design a system meeting certain requirements as regards static and dynamic characteristics.

Chapter I

SIGNALS

1. TYPES OF SIGNALS

Electrical automatic-control systems use electrical signals for the transmission of information. Processing this information consists of the conversion of these signals [from one form to another].

There are two principal types of signal, continuous and discrete, and these determine the method of transmission.

a) Continuous signals

In a continuous signal, the signal amplitude varies continuously (Figure 1.1) and the information transmitted is proportional to the amplitude.

FIGURE 1.1. A continuous signal curve

Until recently the continuous signal has been the most widespread in automatic-control systems. Recently, however, in view of the discrete transmission of information in living organisms and owing to several advantages of discrete automatic-control system (mostly their accuracy and noise discrimination), discrete signals are being used increasingly.

b) Discrete signals

A **discrete** signal differs from a continuous signal in that it varies in steps, and not continuously. For example, the signal whose variation is represented in Figure 1.1 may in reality be following a stepwise curve (Figure 1.2). The information, as in the case of a continuous signal, is given by the ordinates of the curve. The characteristic feature of the discrete signal is that the information has a constant amplitude (i_c) during Δt seconds, and only changes in amplitude at discrete time intervals (each Δt seconds).

Broken signals can also be obtained by converting a continuous curve (the dashed line in Figure 1.2) into a stepwise curve. This conversion is known as the **quantization** of a continuous curve, or analog-to-digital conversion. In Figure 1.2 the signal is **time-quantized**, i.e., the signal changes discretely at definite **time** intervals (Δt).

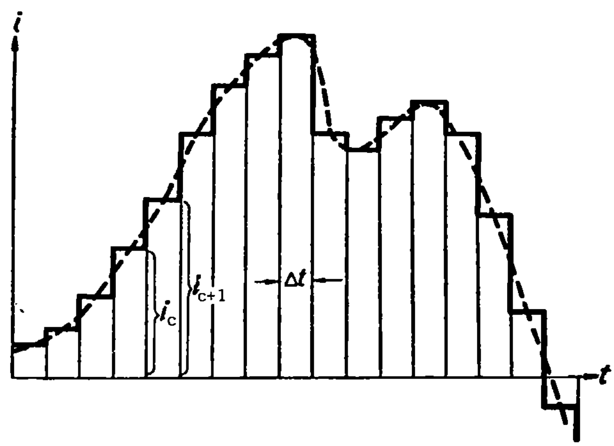

FIGURE 1.2. A discrete current curve.

The higher the frequency of quantization (the smaller Δt), the closer the quantized curve approximates the continuous one. The advantage of discrete over continuous signals is that they are much more easily handled by electronic means. The most important method of discrete representation of signals is **pulse-code modulation**. Here each signal is made to correspond to a code represented by a sequential pulse train, reducing the transmission of information to a transmission of pulse trains.

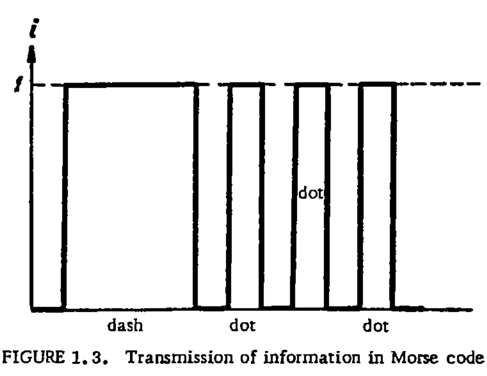

FIGURE 1.3. Transmission of information in Morse code.

The simplest example of pulse-code modulation is the Morse code, where each letter of the alphabet and each number is transmitted as a definite sequence of long (dashes) and short (dots) signals (Figure 1.3).

2. BINARY REPRESENTATION OF NUMBERS

Until recently all computations made in various branches of science and technology used decimal numbers, i.e., combinations of ten elementary digits (symbols):

$$0, 1, 2, 3, 4, 5, 6, 7, 8, 9.$$

This method of number representation is not unique. For example, Roman numerals make use of five symbols:

$$\mathrm{I}, \mathrm{V}, \mathrm{X}, \mathrm{C}, \mathrm{M}^{*},$$

where $\mathrm{I}=1, \mathrm{V}=5, \mathrm{X}=10, \mathrm{C}=100$ and $\mathrm{M}=1000$.

Other number systems are available, the simplest of these being the binary system. This system is based on two symbols only, one and zero.

The decimal number system requires ten different signals (ten different signal levels) to correspond to each digit, whereas the binary system requires only two different signal levels. These signals are formed simply by "one" showing the presence of current flow in the circuit, and "zero" the lack of current flow in the circuit**. The rules of addition and multiplication in the binary system are also very simple. Because of these features, the binary system has become very widespread in computer technology and in cybernetic automatic-control systems.

The control in living organisms is also binary.

A nerve cell (a neuron) is either excited or not. To excite a neuron, the stimulus must attain a certain threshold level. This phenomenon is known in physiology as the "all-or-none" rule.

The shortcomings of the binary system are that the representation of any number requires more digits than in the decimal system, and that a number represented by "1"s and "0"s is unfamiliar. However, these shortcomings are more than compensated by the advantages.

Before proceeding with the discussion of the binary system, let us analyze the representation of numbers in the decimal system. Consider the natural numbers:

$$0, 1, 2, 3, 4, 5 \ldots$$

Only the first ten numbers are represented by corresponding symbols. The transition from the number 9 to the number 10 therefore requires a special procedure. The last of the symbols available (nine) is replaced by the first symbol of the system (zero) and the

* Two additional symbols are sometimes introduced: L = 50 and D = 500.

** Since automatic-control systems generally employ vacuum and semiconductor devices, which are not ideal circuit breakers, the signal "zero" in practice does not correspond to absence of current in the circuit, but rather to the presence of a small current which is readily distinguishable from the current taken to represent "one".

next symbol (one) is written to its left. The number 10 is thus represented with the aid of two digits.

The next number is derived by substituting one for zero (11), the one is then replaced by two (12), etc.

This method of sequential representation of numbers is independent of the number system used. We shall now apply it to the binary system, We have already said that this system makes use of two symbols only. For the sake of convenience the first two digits of the decimal system, 0 and 1, were chosen.

The "zero" and the "one" in the binary system are represented by the same digits as in the decimal system, but a special symbol for representing the number "two" is lacking, and it shall be represented by using the previously described procedure.

The last of the symbols available ("1" in this case) is replaced by the first symbol of the system (zero) and one is written to its left. The number "two" in the binary system is thus represented as "10".

Replacing the zero by the next symbol "1" gives us number "three" ("11").

The procedure for deriving number "four" is the same as used to derive number "two", but now the two symbols (the two ones) are replaced by zeros and a one is written to the left ("100").

Consecutive numbers are written similarly. In conclusion we obtain the following table:

TABLE 1.1

Decimal to binary number conversion

Decimal number	0	1	2	3	4	5	6	7	8	9	10	11
Binary equivalent	0	1	10	11	100	101	110	111	1000	1001	1010	1011

From this table it can be seen that only one binary number, "1", (other than zero) can be represented by a single digit. This can be expressed by the equality $2^0=1$. Two digits will represent three $(2^0+2^1=3)$ binary numbers: "01, 10, 11", three digits, seven $(2^0+2^1+2^2=7)$ binary numbers: "001, 010, 011, 100, 100, 101, 110, 111", etc., and R binary digits [called bits] will represent the following amount of binary numbers:

$$N=2^0+2^1+2^2+\ldots+2^{R-1}=\sum_{k=0}^{R-1} 2^k. \tag{1.1}$$

The series $\Sigma 2^k$ is a geometrical progression with $q=2$ and $a_1=2^0=1$. Using the equations of geometrical progressions, the sum of R terms of this series:

$$\sum_{k=0}^{R-1} 2^k=\frac{a_1(q^R-1)}{q-1}=2^R-1.$$

Hence R bits will represent

$$N=2^R-1 \text{ numbers.} \tag{1.2}$$

From this equation the number of bits required to represent a number N can be calculated:

$$R = \log_2 (N+1), \tag{1.3}$$

where $\log_2 N$ is the logarithm, to the base of 2, of the number N.

Example. Find how many bits are required to represent number thirty-one ($N=31$).

Applying equation (1.3),

$$R = \log_2 (N+1) = \log_2 32 = 5.$$

Hence 5 bits are required to represent the number 31.

Equation (1.3) does not always give an integral number of digits. In this case the result should be rounded to the next greatest number. The resulting (rounded) number of digits (R) is suitable for representing more than N numbers.

Example. Let $N=20$. Then, from equation (1.3),

$$R = \log_2 (N+1) = \log_2 21 = 4.38.$$

Rounding,

$$R = 5.$$

However, from (1.2) it can be seen that five bits will represent

$$N = 2^5 - 1 = 32 - 1 = 31 \text{ numbers}.$$

Five bits are required to represent the first twenty numbers, but they can be used to represent the first thirty-one numbers.

Tables of logarithms to the base of two are generally unavailable in common reference books. However, using well-known conversion formulas, the necessary calculations with any other available tables may be carried out.

The equation $N=2^x$ can be written in the following form:

$$x = \log_2 N,$$
$$x \ln 2 = \ln N,$$
$$x \lg 2 = \lg N,$$

where $\ln N$ = the natural logarithm of the number N;

$\lg N$ = the decimal logarithm of the number N.

Eliminating the unknown x from these equations gives

$$\log_2 N = \frac{\ln N}{\ln 2},$$
$$\log_2 N = \frac{\lg N}{\lg 2}.$$

Inserting the values of $\ln 2$ and $\lg 2$ in this expression gives the conversion formulas

$$\left.\begin{array}{l}\log_2 N = 1.44 \ln N, \\ \log_2 N = 3\,33 \lg N.\end{array}\right\} \qquad (1.4)$$

As regards the representation of fractional numbers, binary systems can be divided into two classes:

a) fixed-point system.

In this case, a binary number, say

$$1011.101,$$

has the binary point at a fixed position, and in general this point is not coded. The integer part of the number is represented with pre-assigned digits, other digits being assigned to the fractional part.

b) floating-point system.

In this system the number N is represented in two parts, the mantissa and the exponent.

For example,

$$101.01 = 0.10101 \cdot 10^{11},$$

where 10 = the number "2" in the binary system;
11 = the number "3" in that system.

In floating-point system the mantissa (10101) and the exponent (11) are coded separately.

The signs of the mantissa and of the exponent are coded as follows:

$$\text{"plus"} \rightarrow 0, \quad \text{"minus"} \rightarrow 1.$$

3. CONVERSION OF BINARY NUMBERS INTO DECIMAL NUMBERS

[To convert a binary number into a decimal number, we must know the decimal equivalent of each unit in the binary number.]

If a binary number has one (zero-order) digit, the corresponding one is equal to the one of the decimal system (Table 1.1). This follows from the equality $1 \cdot 2^0 = 1$. If the zero-order digit is a "zero", we have $0 \cdot 2^0 = 0$. Hence, binary zero is converted to decimal zero. This rule can be extended to any m-th-order digit of the binary number:

$$1 \cdot 2^m = 2^m,$$

$$0 \cdot 2^m = 0. \qquad (1.5)$$

For example, a one in the second-order digit $(m=2)$ is equal to (Table 1.1)

$$1 \cdot 2^2 = 4.$$

This gives us a rule for converting a binary number into its decimal equivalent. The conversion is carried out by inserting zeros and ones as coefficients in the terms of the series (1.1) in accordance with the digits of the given binary number.

Example. Find the decimal equivalent of the binary number 100110.

We write six terms in series (1.1), since the given binary number has six digits (the zero-order digit, the first-order digit, ..., the fifth-order digit):

$$2^5+2^4+2^3+2^2+2^1+2^0.$$

Zeros are then written in as coefficient preceding the zero-, third-, and fourth-order digits of the number 100110, and ones, before the remaining digits, giving

$$N=1\cdot 2^5+0\cdot 2^4+0\cdot 2^3+1\cdot 2^2+1\cdot 2^1+0\cdot 2^0=38.$$

4. CONVERSION OF DECIMAL NUMBERS INTO BINARY NUMBERS

Any decimal number N can be written, as in (1.1), in the following form:

$$N=2^{R-1}+a_1\cdot 2^{R-2}+\ldots+a_{R-3}\cdot 2^2+a_{R-2}\cdot 2^1+a_{R-1}\cdot 2^0, \qquad (1.6)$$

where R = the number of digits required for the binary representation of the number (see equation (1.3));

a_1—a_{R-1} = coefficients equal to 0 or 1.

In the first method of conversion, the given decimal number is expanded into a series of the form $\Sigma\, 2^m$ and the exponents m of this expansion are established. The equivalent binary number is obtained by writing ones at the positions corresponding to these exponents, and zeros in the remaining positions.

Example. Find the binary equivalent of the number 43.

Expanding the number 43 into a series in powers of 2 gives

$$43=32+8+2+1=2^5+2^3+2^1+2^0.$$

The first, second, fourth, and sixth digits (or, in other words, the zero-order, first-order, third-order, and fifth-order digits) of the binary number are ones, and the third and the fifth (or the second- and the fourth-order) digits are zeros. Thus

$$43\rightarrow 101011.$$

Another, simpler method of decimal-to-binary conversion is as follows: both sides of equation (1.6) are divided by 2. Seeing that $2^0=1$, we obtain

$$\frac{N}{2}=2^{R-2}+a_1\cdot 2^{R-3}+\ldots+a_{R-3}\,2^1+a_{R-2}+\frac{a_{R-1}}{2}. \qquad (1.7)$$

This division of N by 2 produces a quotient $(2^{R-2}+a_1 \cdot 2^{R-3}+\ldots+$ $+a_{R-3}2^1+a_{R-2})$ and a residue (a_{R-1}). If N is even, the residue $a_{R-1}=0$; if N is odd, $a_{R-1}=1$.

Dividing $\frac{N}{2}$ again by 2 (and omitting the term $\frac{a_{R-1}}{2}$ gives the residue a_{R-2} [which is equal to zero if $\frac{N}{2}$ is even, and equal to one if $\frac{N}{2}$ is odd, and so on].

The rule for decimal-to-binary conversion is as follows:

The given decimal number N is divided by 2. If N is an even number, the zero-order digit of the binary number is a zero. If N is an odd number, a residue of one is obtained, and the zero-order digit is a one. The quotient $\frac{N}{2}$ is then divided by two (omitting the previous residue). If the division gives no residue, the first-order digit of the binary number is a zero; if, however, a residue of one is obtained, the first-order digit is a one. This process is continued until the final quotient is zero.

In practice this method is applied as follows: the given decimal number is successively divided by 2, and the residue of each division (zero or one) is written down in a column. This column represents the equivalent binary number.

Example. Find the binary equivalent of the number 69. Divide 69 by 2. The quotient is 34 with a residue of 1. The residue is written down in the right-hand column, and 34 is again divided by 2, etc. This process is continued until the quotient is zero. This conversion is represented in the following table:

69

Quotient	Residue
34	1
17	0
8	1
4	0
2	0
1	0
0	1

Hence,

$$69 \rightarrow 1000101.$$

5. QUANTIZATION OF CONTINUOUS SIGNALS

Industrial installations often deal with continuous processes. The signals arising from any variation in the parameters of these processes are therefore also continuous. [They are called analogs of the measured parameters.]

It would appear that, when continuous signals are converted into discrete signals (i.e., when the signals are *quantized*), some of the information

contained in continuous signal would be partially lost. V. A. Kotel'nikov has proved that this is not so, and that continuous signals can be quantized without any loss of information.

The proof of Kotel'nikov's theorem is based on the following. First, any signal transmitted in a real channel has a limited frequency spectrum, because when passing through a real apparatus, say through an electric-motor winding, the signal is smoothed out and the higher harmonics are attenuated so as to become negligible. Second, the signal is invariably subject to distortion from noise, and the transmission of this signal with an error smaller than the noise magnitude therefore corresponds in effect to transmission with no error at all. A continuous function with a frequency spectrum whose upper limit is f_{max} can therefore be exactly represented by a finite number of its values measured at time intervals not exceeding (Figure 1.2)

$$\Delta t = \frac{1}{2 f_{max}}. \qquad 1.8$$

Thus, it is possible to convert a continuous signal into a discrete signal without loss of information.

6. ANALOG-TO-DIGITAL CONVERSION

Various transducers developing voltage analogs of various parameters may be used. These transducers measure displacement, temperature, pressure, and a multiplicity of other parameters which vary continuously with time.

To apply this information in digital automatic-control equipment, the analog signal must be quantized.

Different types of electronic or electromechanical equipment are available for this conversion. Electronic converters are capable of up to a million binary conversions per second. They may have no moving contacts and are much more reliable in service than mechanical devices. It should be kept in mind, however, that some electromechanical converters are simpler and cheaper than the electronic ones. The choice of the converter therefore depends on the actual problem encountered.

Two main techniques are currently used for analog-to-digital conversion. These are the time-base encoding method and the weighted-voltage comparison method, whose essentials we shall now briefly consider.

a) Time-base encoding method

In time-base encoding a sawtooth-voltage generator is used. This generator sets up periodic voltage (Figure 1.4), which increases proportionally with time during the trace interval and drops to zero during the flyback interval.

The continuous voltage $u=f(t)$ to be quantized is superimposed on the sawtooth voltage.

At points B_1, B_2, B_3, etc., these voltages are equal. The voltage at

point B_1 is therefore given by

$$u_1 = t_1 \cdot \operatorname{tg} \alpha.$$

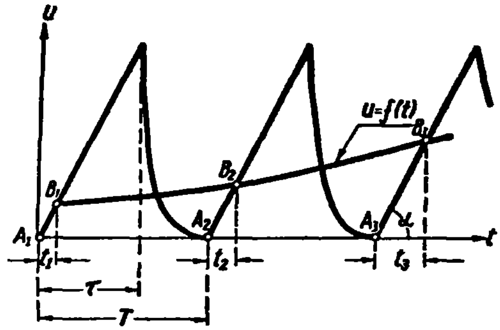

FIGURE 1.4. Comparison of two voltages.

Since the slope (α) of the sawtooth voltage is constant, the voltage at point B_1, taken to an appropriate scale, is equal to the time interval t_1. The continuous voltage can be quantized in the following manner.

A clock-pulse generator (Figure 1.5) sends pulses simultaneously to a counter and scaler. The scaler produces an output signal after receiving n pulses, i.e., at time intervals

$$T = \frac{1}{f_p} n,$$

where f_p is the clock-pulse generator frequency. The output signal from the scaler synchronizes the pulse generator with the counter at points A (Figure 1.4)

The input to the voltage comparator consists of the sawtooth voltage and of the voltage to be quantized (u). When these voltages are equal (points B in Figure 1.4), the comparator sends an inhibit pulse to the gate, inhibiting any further pulses from the pulse generator to the counter. The number of pulses transmitted to the counter while the gate is open is proportional to the voltage measured at point B. After counting these pulses, the counter generates a discrete output voltage u. The counter used determines the coding system of this output voltage. If a binary counter is used, the output voltage is binary-coded. Having generated an output signal, the counter clears itself to zero, making it ready for counting the voltage (u) in the next interval.

The advantage of the *time-base encoder is the simplicity of the circuit* and equipment required. Its shortcoming is the relatively low accuracy due to the inaccuracy in the count of the unequal time intervals and uncertainty of count.

The low accuracy arises from any nonlinearity in the sweep voltage.

Moreover, the origin of the pulse count (points A) is determined with insufficient accuracy and requires a special circuit for its determination.

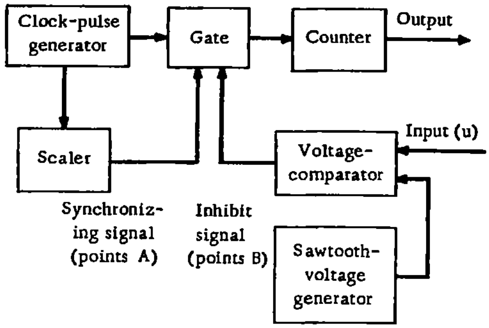

FIGURE 1.5. Time-base encoding system.

Uncertainty of count arises because the high-speed counter may register extra pulses during the application of the inhibit signal.

b) Weighted-voltage comparison method

This method is based on the comparison of instantaneous values of the continuous voltage (u) to be converted, with several standard voltages. The standard voltages are of the form $u_{st} = 2^k$, where k is a positive integer. The standard voltages used are :

$$u_{st_0} = 1\,\text{v};\ u_{st_1} = 2\,\text{v};\ u_{st_2} = 4\,\text{v};\ u_{st_3} = 8\,\text{v}\ldots$$

The comparison of the analog voltage with the standard voltages follows the logical rules below.

The highest standard voltage (u_{st_k}) is taken and a comparison is made.

1. If $u \geqslant u_{st_k}$, an output of one is produced for the k-th-order digit, and the sum of two standard voltages

$$u_{st_k} + u_{st_{k-1}}$$

is next used for comparison.

2. If $u < u_{st_k}$, a zero signal is produced in the higer k-th-order digit and the next smaller standard voltage $\left(u_{st_{k-1}}\right)$ is taken for comparison.

3. Depending on the results in articles 1 and 2 above, a similar

comparison is made following one of the conditions below:

$$u > u_{st_{k-1}} \text{ or } u > u_{st_k} + u_{st_{k-1}};$$

if either inequality is satisfied, the (k-1)-th-order digit receives a unit signal, otherwise a zero signal is produced.

4. Similar comparisons are made with sums of all the lower-order digits.

Example. Convert the voltage $u=58$ v into binary form.

The standard voltages available in the converter are 1, 2, 4, 8, 16, 32, and 64 v.

The sequence of operation of a binary weighing encoder is shown in Table 1.2.

TABLE 1.2

Logical operation of binary-weighing encoder

No.	Position	Comparison	Result of comparison	Output
1	6	$58 > 64$	No	0
2	5	$58 > 32$	Yes	1
3	4	$58 > 32+16$	Yes	1
4	3	$58 > 32+16+8$	Yes	1
5	2	$58 > 32+16+8+4$	No	0
6	1	$58 > 32+16+8+2$	Yes	1
7	0	$58 > 32+16+8+2+1$	No	0

We thus have the conversion

$$58 \text{ v} \rightarrow 0111010.$$

Since the pulses produced by the encoder are in succession, this system is called serial.

By using a beam-switch tube in this encoder, a conversion frequency of up to 100 kc can be achieved. The conversion accuracy achieved by this method is higher than that inherent in the time-base encoding method, since it is limited only by the accuracy of the standard voltages and the comparison circuit.

A disadvantage of this method is the need for an involved logical circuit.

7. DIGITAL-TO-ANALOG CONVERSION

It should be kept in mind that the discrete signal represents the value of the variable parameter only at those instants when the measurements were made. Consequently, the value of the continuous signal can be determined only at these definite points. This is similar to plotting a curve from several

points, where no information is available on the behavior of the curve between each two points.

The conversion of a discrete signal to a continuous signal usually consists of converting a binary (and sometimes a decimal) code into a continuous voltage, whose magnitude is an analog of the number specified by the code.

The most common method of converting this code into its analog equivalent is the w e i g h i n g m e t h o d, so called, because it involves weighing all the ones of the binary number.

Take for example the number 111. From left to right the digits "weigh" 4 v, 2 v and 1 v. Adding them together gives 7 volts.

A circuit using this method for the conversion of a four-digit binary number is shown in Figure 1.6

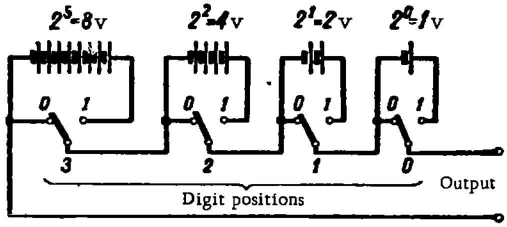

FIGURE 1.6. A circuit for a binary weighing decoder.

If R standard voltage sources are provided in this circuit, it can be used to convert any R-digit number.

The operation of this circuit is based on the switching of the appropriate contact by the appropriate pulse. When no pulse is received (i.e., the digit is zero), the corresponding contact remains to the left. If a one is received the contact of the corresponding digital position is switched over to the right, and the standard voltage representing the weight of the digit in question is delivered to the output.

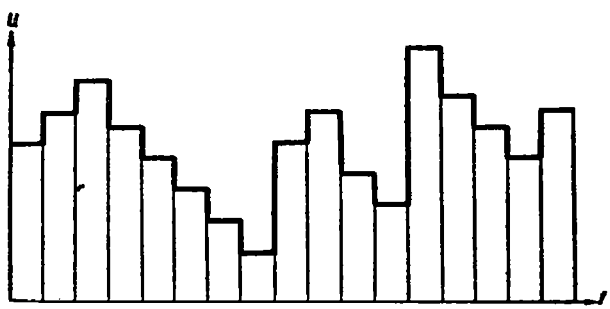

FIGURE 1.7. Converter output voltage.

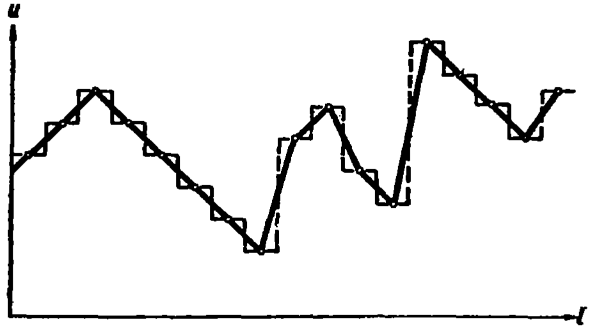

FIGURE 1.8. Smoothed converter output signal.

For example, if the number 1011 is received, the contacts are switched over to the right in positions 3, 1, and 0, and voltages of 8, 2, and 1 v, respectively, are delivered to the output.

When a conversion of a decimal number is required, the standard voltages should be 1, 10, 100 v, etc.

Electronic switches in the circuit of Figure 1.6 can convert binary-coded information at a rate of up to 1 million pulses per second.

The voltage output from the converter is plotted in Figure 1.7. If suitable filters are provided at the output, a smoother voltage curve is obtained (Figure 1.8). The output voltage is exact only at those points where conversion was made. Thus, the higher the frequency of conversion, the greater the accuracy of the voltage curve obtained. As previously indicated (Kotel'nikov's theorem), the appropriate selection of conversion frequency will ensure a continuous output voltage without loss of information, i.e., with an accuracy exceeding the errors of measurements.

Chapter II

ELECTRONIC LOGICAL ELEMENTS

1. FUNDAMENTAL CONCEPTS OF THE ALGEBRA OF LOGIC

Logic is the science of the forms and laws of reasoning. It is divided into two parts, dialectic and formal. Formal logic describes the forms of reasoning as constant and immutable. Dialectic logic is concerned with the evolution of the forms and laws of reasoning, and it is this factor of evolution which distinguishes it from formal logic. Evolution, in turn, is determined by the interactions inherent between each and every phenomenon. Dialectic logic is at present in its early stages of development. Formal logic, however, is a science which has been studied in considerable detail.

A branch of formal logic which is readily adaptable to the requirements of mathematics is mathematical logic. One branch of mathematical logic is the algebra of logic or Boolean algebra. The latter name derives from G. Boole, one of the founders of the algebra of logic. Boole was the first to develop the calculus of propositions which operates with logical statements much in the same way as conventional mathematics operates with algebraic symbols.

Let A stand for a statement, which in mathematics is generally called a proposition. Any proposition is either true (correct), or false (incorrect). The truth of a proposition shall be denoted by the word "YES", and its falseness by the word "NO". If the proposition is transmitted using binary representation, its truth will be represented by the signal "1", and its falsness by the signal "0". For example, let A represent the proposition "The coil is energized". If the proposition is true, $A = 1$, and if false, $A=0$.

Boolean algebra enables us to solve two kinds of problems:

a) problems of analysis, where it is required to describe the logical operations of a given electronic or electromechanical circuit;

b) problems of design, where it is required to design an electronic or electromechanical circuit to perform given logical requirements.

Any compound proposition consists of simple propositions. Conversely, simple propositions can be used to construct a compound proposition. Therefore, using simple electronic or electromechanical elements, complicated logical propositions may be designed.

a) Negation of a proposition

This operation is denoted symbolically by $C=\bar{A}$ and reads "Proposition C is true when proposition A is not true (false)". Since a proposition is

qualified by the symbols "1" or "0", the negation of a proposition is represented by the following equalities:

$$\bar{0}=1,$$
$$\bar{1}=0, \qquad (2.1)$$

i.e., the negation of zero is one, and the negation of one is zero.

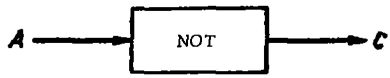

FIGURE 2.1. Symbol representing a NOT gate.

In electronic circuits, the logic element performing negation is denoted (Figure 2.1) by a rectangle with the word "NOT" inscribed, [and is called a NOT gate]. An arrow pointing toward the rectangle indicates the input of the gate, and another arrow pointing out, indicates the output. An output signal (1) is produced only when no input signal (0) is received. When an input signal is received by the gate, it produces no output signal.

b) Conjunction of two propositions

The conjunction of two propositions is denoted symbolically by $C=A\wedge B$ or $C=A\cdot B$ and reads "Proposition C is true if propositions A and B are true". The "dot" in the second symbolical expression is used because conjunction defines a table for the logical multiplication of two numbers (for a binary system):

$$0\cdot 0=0,$$
$$0\cdot 1=0,$$
$$1\cdot 0=0,$$
$$1\cdot 1=1. \qquad (2.2)$$

The logic element used to perform conjunction is denoted by a rectangle (Figure 2.2) with the word "AND" inscribed [and is called an AND gate]. Since this element compares two propositions A and B, it has two inputs. An output signal ($C=1$) is produced when propositions A and B are true, i.e., when $A=1$ AND $B=1$. In other words, an output signal is produced when signals are received at both inputs. If the element performs the conjunction of n propositions, it has n inputs.

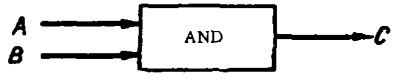

FIGURE 2.2. Symbol representing an AND gate.

c) Disjunction of two propositions

The disjunction of propositions is denoted either by $C=A\nabla B$ or $C=A+B$, and reads "Proposition C is true if proposition A OR proposition B is true".

Disjunction defines the logical addition of two binary numbers:

$$\begin{aligned} 0+0&=0, \\ 0+1&=1, \\ 1+0&=1, \\ 1+1&=1. \end{aligned} \tag{2.3}$$

A logic element performing the disjunction of two propositions is denoted by a rectangle with the word "OR" inscribed (Figure 2.3) [and is called an OR gate].

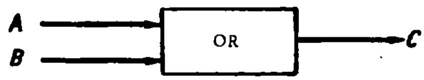

FIGURE 2.3. Symbol representing an OR gate.

An output signal is produced when a signal is received at either input A OR B. In the disjunction of n propositions, the element has n inputs.

More complicated propositions can be expressed in terms of negation, conjunction, and disjunction. For example, the proposition $C=AB$, where $B=D+E$, is written.

$$C=A(D+E)$$

Thus, using electrical circuits, a complicated logical function can be performed, using the simple NOT, AND, and OR gates.

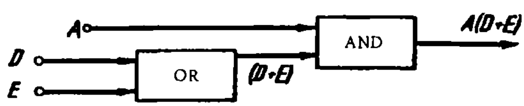

FIGURE 2.4. Symbolic diagram of the logical operation $A(D+E)$.

The logical function in the previous example is performed using two gates (Figure 2.4). The OR gate performs the disjunction of propositions $(D+E)$, and the AND gate performs the conjunction $A(D+E)$.

Note that the operations of negation, conjunction, and disjunction are not independent. They are related by the following equations:

$$\begin{aligned} A+B&=\overline{\bar{A}\bar{B}}; \\ AB&=\overline{\bar{A}+\bar{B}}. \end{aligned} \tag{2.4}$$

A bar over the two letters indicates the negation of the logical product or logical sum.

The proof of equation (2.4) is fairly simple. Symbols A and B, unlike their use in conventional algebra, do not assume any arbitrary value, but are either 1 or 0. Therefore, considering (2.4) for the four possible cases

$A=$	0	0	1	1
$B=$	0	1	0	1

(2.5)

and applying the rules of logical multiplication (equation (2.2)) and logical addition (equation (2.3)), we easily arrive at their validity.

The types of electronic or electromechanical elements required to perform a complicated logical function can be reduced to a single type. Consider an AND-NOT element [called a NAND gate] (this element is often denoted by the letter L), which performs the following logical proposition (Sheffer's stroke function):

$$C=\overline{AB}.$$

Given NAND gates, circuits may be designed to perform negation, disjunction, and conjunction (Figure 2.5). The proof of this is evident when we consider all the possible (2.5) values of propositions

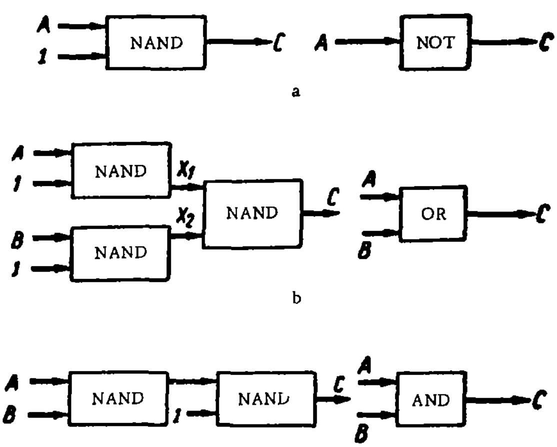

FIGURE 2.5. (a) NOT circuit, (b) OR circuit, (c) AND circuit designed using NAND gates.

As an example of the proof, consider the OR circuit (Figure 2.5, b). Taking X_1 and X_2 as the outputs of the first two NAND gates, we obtain

$$X_1=\overline{A\cdot 1};\ X_2=\overline{B\cdot 1};\ C=\overline{X_1X_2}.$$

The results of these operations for all the possible values of A and B are arranged in Table 2.1.

TABLE 2.1
Logical operations performed by the circuit in Figure 2.5,b.

A	B	X_1	X_2	C
0	0	1	1	0
0	1	1	0	1
1	0	0	1	1
1	1	0	0	1

Comparing the results obtained with the rules of logical addition (equation (2.3)), we observe that the circuit in Figure 2.5, b performs the OR function.

Another element has recently been introduced which can be used to produce all the other logical operations. This element, symbolically denoted by a rectangle with an inscription OR-NOT (Figure 2.6) [called a NOR gate] performs the following logical operation:

$$C=\overline{A+B}.$$

Using the procedure discussed for the circuits in Figure 2.5, it can be shown that the NOR gate can produce the NOT, OR, and AND functions (Figure 2.7).

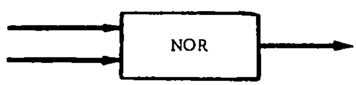

FIGURE 2.6. Symbol representing a NOR gate.

NAND and NOR gates are generally not used as substitutes for the other three types of logic elements. In some cases, however, it is possible to realize complex functions more simply by the use of these special gates rather than with NOT, OR and AND elements.

Boolean algebra obeys the same rules of addition and multiplication as conventional mathematics. Among the basic rules we have:

$$A+B=B+A; \qquad AB=BA;$$
$$(A+B)+C=A+(B+C); \quad (AB)C=A(BC);$$
$$A(B+C)=AB+AC.$$

Since the symbols (A, B , etc.) in Boolean algebra may be equal to 0 and 1 only, it should be kept in mind that, when manipulating with these formulas, equalities not characteristic of elementary algebra may be obtained. For example:

$$\begin{aligned} &A+A=A;\ AA=A; \\ &A+\bar{A}=1;\ \ A\bar{A}=0, \\ &1+A=1;\ A+AB=A. \end{aligned} \tag{2.6}$$

Equalities (2.6) can be proven by using the same technique as used to prove equalities (2.4).

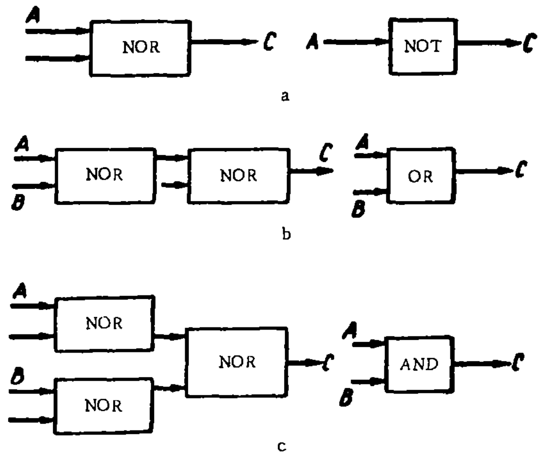

FIGURE 2. 7. NOR gates producing the NOT, OR, and AND functions.

It must be remembered that the comparison between logical and conventional algebraic addition and multiplication is not perfect. For example, in conventional algebra $A+A=2A$, whereas in Boolean algebra (see (2.6)) $A+A=A$. Moreover, some operations of Boolean algebra (e.g., negation) have no direct comparison in conventional algebra.

Using the rules of Boolean algebra, various complicated logical expressions may be simplified.

Example. Design a circuit to perform the following function:

$$C=AB(AB+B).$$

The function specified can be performed using three gates (Figure 2.8, a). It can, however, be substantially simplified. Indeed,

$$C=AB(AB+B)=AABB+ABB.$$

But from (2.6)

$$AA=A;\ BB=B.$$

Hence

$$C=AB+AB.$$

Moreover, from (2.6), $D+D=D$. Hence, setting $D=AB$, we obtain

$$C=AB.$$

The function specified can thus be performed using one gate only (Figure 2.8, b).

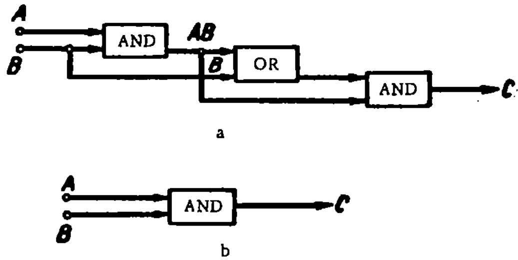

FIGURE 2. 8. Two circuits which perform the same logical function.

2. BINARY ARITHMETIC

Systems may be devised using electronic logic elements, which will add, subtract, multiply, and divide, according to the rules of logical multiplication and addition.

a) Binary addition

Binary numbers are added according to the rules of logical addition (2.3). In this addition a cary of one is added to the next highest position when two ones are added. We therefore have the following rules for the addition of one-bit numbers:

$$
\begin{aligned}
&0+0=0,\\
&0+1=1,\\
&1+0=1,\\
&1+1=10
\end{aligned}
\qquad (2.7)
$$

These operations can be performed by the circuit shown in Figure 2.9.

The carry signal (C) is produced, as can be seen from the fourth equality in (2.7), only when the two addends A and B) are ones. This can be written in the form

$$C=A \text{ and } B$$

or

$$C=AB.$$

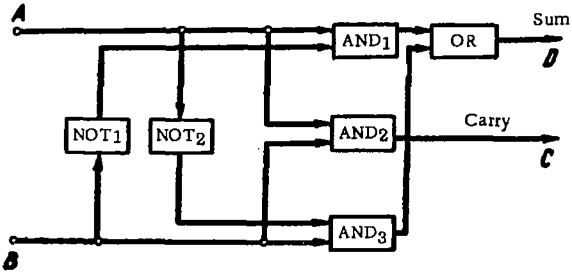

FIGURE 2. 9. Adder circuit for two binary numbers.

The carry signal is therefore the output of the AND_2 gate whose inputs are the signals A and B (Figure 2.9).

The sum signal is obtained (see (2. 7)) in two cases, when

$$(A=1,\ B=0) \text{ or } (A=0,\ B=1).$$

Using Boolean terminology, we may write these conditions in the form:

[A and (not B)] or [(no A) and B]

In symbolic notation, the equivalent is:

$$A\bar{B} \text{ or } \bar{A}B.$$

Finally, this function can be written:

$$D=A\bar{B}+\bar{A}B, \tag{2.8}$$

where D is the sum signal.

This addition can be performed (Figure 2. 9) using five gates, namely NOT_1 , NOT_2 , AND_1, AND_3 , and OR. Equation (2. 8), however, does not contain the optimum number of gates for deriving the sum D. This is because the previously formed signal $C=AB$ has not been utilized. This point will now be considered.

Seeing that $A\bar{A}=0$ and $B\bar{B}=0$ (2.6), we may write (2.8) as

$$D=A\bar{B}+\bar{A}B=A\bar{A}+A\bar{B}+\bar{A}B+B\bar{B}.$$

After simplifying, we obtain

$$D=(A+B)(\bar{A}+\bar{B}).$$

It follows from the second equation of (2.4) that

$$\overline{AB}=\bar{A}+\bar{B}.$$

Hence

$$D=(A+B)\overline{AB};$$

but

$$AB=C.$$

Giving finally

$$D=(A+B)\bar{C}.$$

This expression is simpler than that of (2.8). Thus, the addition is performed (Figure 2.10) using four (and not six, as in the previous case) gates.

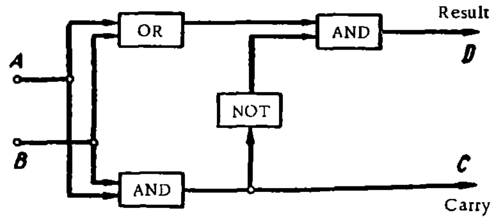

FIGURE 2.10. Simplified adder circuit.

The circuits in Figures 2.9 and 2.10 are used only in addition of one-bit numbers. To add two R-bit numbers, we must have an adder consisting of $[R]$ half-adders (HA). An adder for three-bit numbers is shown in Figure 2.11. Two three-bit numbers ($A_2A_1A_0$ and $B_2B_1B_0$) constitute the input of this device. The output is a four-bit number ($D_3D_2D_1D_0$).

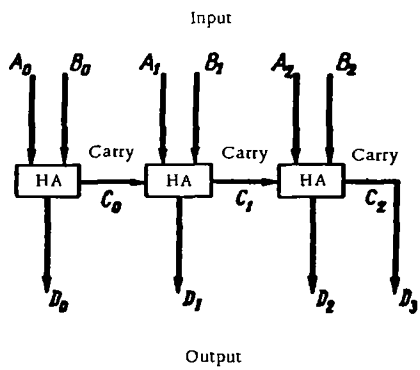

FIGURE 2.11. Adder for three-bit numbers.

From this diagram it can be seen that each half-adder (HA) should be so designed that it adds three one-bit numbers, namely A, B, and the carry from the lower position (C). The circuit of one of these adders, assembled from NOT, OR and AND gates is shown in Figure 2.12.

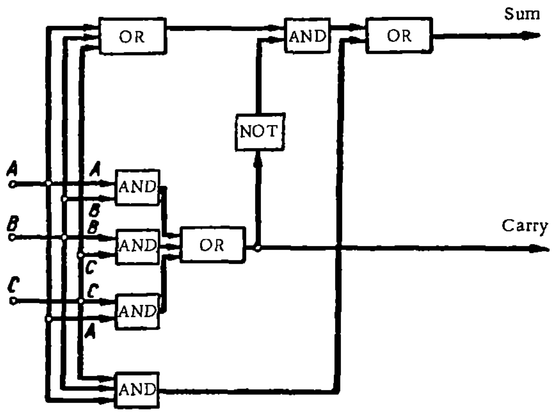

FIGURE 2.12. Adder for three one-bit numbers.

b) Binary subtraction

Binary numbers can be subtracted by two methods. First, a unit can be designed which, subtracting a smaller number from a greater number, will subtract directly the bits of the same position according to the following rules:

$$\begin{aligned} &0-0=0,\\ &1-0=1,\\ &0-1=1(1),\\ &1-1=0, \end{aligned} \tag{2.9}$$

where (1) indicates a carry (add 1) to the next highest position of the subtrahend.

Using these rules, a logical circuit for the subtraction of two one-bit numbers, can be easily designed.

Let the diminued be A and the subtrahend B. The carry signal (C) is produced when

$$C = (\text{not } A) \text{ and } B$$

or

$$C = \bar{A}B.$$

This logical function is produced by the two gates, NOT_2 and AND_2 (Figure 2.13). A one is obtained after subtraction if $A=1$ and $B=0$ or if $A=0$ and $B=1$. This function, as we have shown in the discussion of binary addition, is written as

$$D=A\bar{B}+\bar{A}B.$$

The subtraction in Figure 2.13 is performed by five gates, two of which are used for producing the carry signal (C).

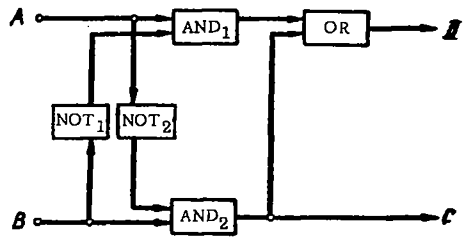

FIGURE 2.13. Subtractor circuit for two one-bit numbers.

Comparing the adder (Figure 2.9) and subtractor (Figure 2.13) circuits for one-bit numbers we see that they differ only in the logic of carry determination. It is often convenient to use one circuit for both addition and subtraction. In this case the subtraction is reduced to addition by a special technique, now to be described.

Consider the subtraction of two R-bit numbers:

$$A-B.$$

The greater of these two numbers is *scaled*, i.e., it is written in the form

$$A=a10^n,$$

where a = the *mantissa* of the binary number. It is smaller than one, and has a one in the most significant position immediately following the binary point;

n = the *scale factor* (exponent) of the number;

10 = "two" in binary representation.

Example. Scale the binary number

$$A=101.10.$$

We write this number so that its integral part is zero and it has 1 in the most significant position after the binary point:

$$A=0.10110\cdot 10^{11}.$$

Next, we write the smaller of the two numbers (B) in the form

$$B=\beta 10^n,$$

where n is the scale factor of the scaled number A.

We can thus write the equation

$$A-B=(\alpha-\beta)\ 10^n=(\alpha-\beta+1-1)\ 10^n. \tag{2.10}$$

We now introduce a new number (δ), defined as

$$\delta=1-\beta.$$

From this equality,

$$\beta+\delta=1, \tag{2.11}$$

i.e., δ is the one's complement of β

This complement is derived in the binary system by substituting 1's for all the 0's (to the right of the binary point) and 0's for all the 1's, and then adding 1 to the least significant digit. The validity of this procedure is easily proved by adding the complement δ derived in this way to the initial number β and seeing that the result (see (2.11)) is indeed 1.

Substituting β from (2.11) into (2.10) gives

$$A-B=(\alpha+\delta-1)\,10^n. \tag{2.12}$$

Comparing (2.10) and (2.12), we may write

$$\alpha-\beta=\alpha+\delta-1.$$

Thus, to subtract two numbers, we need only to add the one's complement of the subtrahend to the diminued and to subtract 1 from the result. The subtraction of two numbers is thus reduced to addition.

Let us consider some examples of subtraction of two numbers using the rules of logical addition (2.7).

Example 1. $A=1000$ (eight), $B=101$ (five).

Find $A-B$ (diminuend greater than subtrahend).

1. Scale the greater number:

$$A=0.1\cdot 10^{100}, \text{ i.e. } \alpha=0.1000.$$

2. Equate the scale factor of the numbers A and B ($n=100$):

$$B=0.0101\cdot 10^{100}, \text{ i.e. } \beta=0.0101.$$

3. Find one's complement of β:

$$\begin{array}{r} \beta=0.0101 \\ \downarrow\downarrow\downarrow\downarrow \\ 0.1010 \\ +\quad\ \ 1 \\ \hline \delta=0.1011 \end{array}$$

4. Add $\alpha+\delta$:

$$\begin{array}{r} + \begin{array}{r} 0.1000 \\ 0.1011 \end{array} \\ \hline 1.0011 \end{array}$$

5. Subtract 1 from the sum $(\alpha+\delta-1)$:

$$1.0011-1=0.0011.$$

Thus,

$$A-B=0.0011 \cdot 10^{100}=11 \text{ (three)}.$$

Example 2. $A=110$ (six), $B=1011$ (eleven).
Find $A-B$ (diminued smaller than subtrahend).
1. Scale the greater number :

$$B=0.1011 \cdot 10^{100}, \text{ i. e. } \beta=0.1011.$$

2. Equate the scale factor ($n=100$) of the numbers A and B:

$$A=0.011 \cdot 10^{100}, \text{ i. e. } \alpha=0.0110.$$

3. Find the one's complement of β :

$$\begin{array}{r} \beta=0.1011 \\ \downarrow\downarrow\downarrow\downarrow \\ + \begin{array}{r} 0.0100 \\ 1 \end{array} \\ \hline \delta=0.0101 \end{array}$$

4. Add $\alpha+\delta$:

$$\begin{array}{r} + \begin{array}{r} 0.0110 \\ 0.0101 \end{array} \\ \hline 0.1011 \end{array}$$

5. Subtract 1 from the sum :

$$\alpha+\delta-1=0.1011-1.$$

Since the result of this subtraction is a negative number, we write it in the following form:

$$\alpha+\delta-1=-(1-0.1011).$$

As we have previously shown, the subtraction of the two numbers in parentheses can be reduced to addition. To accomplish this, the one's complement of the subtrahend (0.1011) must be added to the diminuend (1) and 1 must be subtracted from the result.

Hence, $\alpha+\delta-1$ is equal to the one's complement of 0.1011 taken with the minus sign.

6. Find the one's complement of 0.1011:

$$\begin{array}{r} 0.1011 \\ \downarrow\downarrow\downarrow\downarrow \\ 0.0100 \\ +\quad\quad 1 \\ \hline 0.0101 \end{array}$$

Therefore,

$$A-B=-0{,}0101\cdot 10^{100}=-101 \text{ (minus five)}.$$

c) Binary multiplication

One-bit numbers are multiplied according to the rules of logical multiplication (2.2). It follows from these rules that multiplication of one-bit numbers can be performed using the logical element AND.

R-bit numbers are multiplied as in ordinary algebra, i.e., by a combination of addition and shift.

Example. Multiply 101 (five) and 1011 (eleven).

The multiplication is performed as follows:

$$\begin{array}{r} \times\quad 1011 \\ 101 \\ \hline +\quad 1011 \\ 1011 \\ \hline 110111 \end{array} \text{ (fifty-five).}$$

d) Binary division

Since one-bit numbers assume two values only, 0 and 1, their rules of division are very simple:

$$\begin{aligned} 0:1&=0,\\ 1:1&=1. \end{aligned} \tag{2.13}$$

R-bit numbers are divided as in ordinary algebra, and division is in effect reduced to subtraction.

Example. Divide 110111 (fifty-five) by 101 (five).

The division is performed as follows:

```
_ 110111 | 101
  101    |------
  ------ | 1011 (eleven)
_ 00111
    101
   ----
_  0101
    101
    ---
    000
```

3. ELECTRONIC LOGIC CIRCUITS

We have previously shown that the basic electronic logic elements used in the design of logic circuits are the NOT, OR, and AND gates. These gates can be assembled of valves, semiconductors, ferrites, superconductor devices, parametric resonators, etc. Because of the limited scope of this book, we shall consider only the more common gates using vacuum tubes and semiconductor devices.

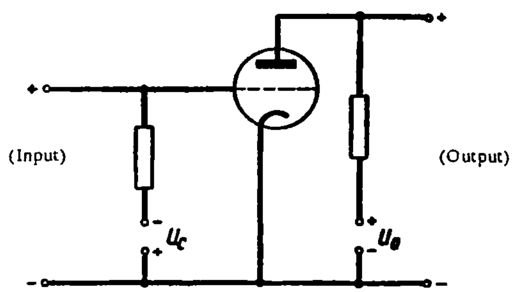

FIGURE 2.14. Triode NOT gate circuit.

a) NOT gate circuits

A triode NOT circuit (often called inverter) is shown in Figure 2.14. In this circuit, a cut-off voltage u_c is applied to the grid of the tube. When no input signal is received, the valve is cut off and the output u_a is taken as 1. To simplify the discussion, we shall assume the internal resistance of the

valve to be zero. If a positive voltage greater than u_c is applied to input (A), the tube conducts heavily and the output voltage (B) drops to zero*. The circuit in Figure 2.14 thus performs the negation

$$B=\bar{A}.$$

A NOT gate can also be assembled using a transistor (Figure 2.15). In this circuit the transistor is cut off when no input signal is received. The corresponding output is then u_{ec}. When a pulse is applied to the input, the transistor saturates and the output voltage drops to zero (the internal resistance of the saturated transistor is neglected).

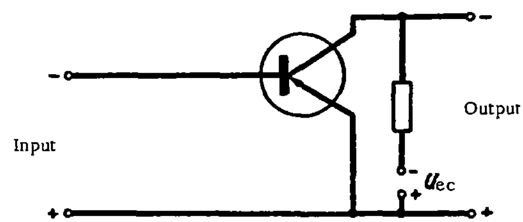

FIGURE 2.15. Transistor NOT gate circuit.

A diode NOT gate is shown in Figure 2.16. If $R_1=R_2$, the output voltage is equal to $\frac{1}{2}u$ when the input voltage is zero. When a voltage greater than u is applied at the input, the diode is cut off; the voltage across the resistance R_1 is opposed to u, so that the output voltage drops to zero.

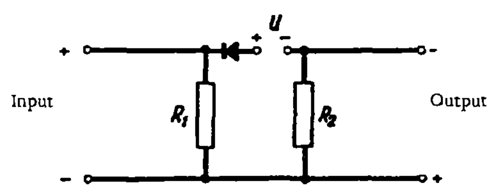

FIGURE 2.16. Diode NOT gate circuit.

b) AND gate circuits

This gate has several inputs whose number is equal to the number of signals required to produce an output signal.

First consider a pentode AND circuit (Figure 2.17). An output signal in this circuit is produced only when positive voltages are received at both inputs. Diodes are provided to prevent a positive voltage applied to one grid

* Actually the tube has a finite internal resistance and the output voltage does not drop to zero. This voltage, however, can still be considered as the zero signal, as its value is only one-fifth to one-tenth of the value of u_a, when it has the value of 1.

from being impressed on the second grid. If the internal impedance of the voltage source u_c is low, these diodes may be dispensed with.

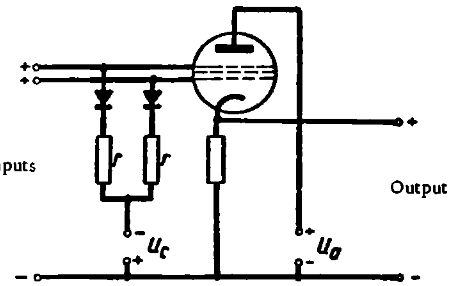

FIGURE 2. 17. Pentode AND gate.

An AND gate using transistors is shown in Figure 2. 18. An output signal is produced when the two transistors are saturated. If an n-input AND gate is required, n-transistors in series must be provided.

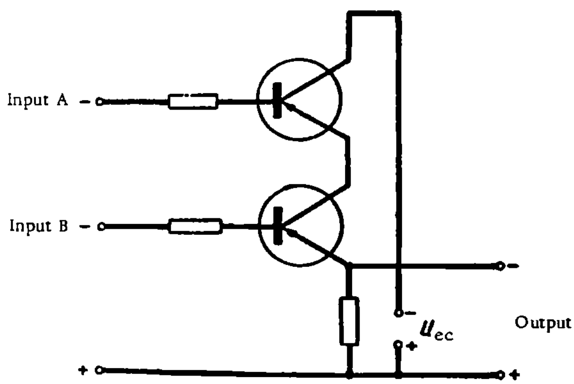

FIGURE 2. 18. Transistor AND gate circuit.

An AND gate can also be designed using diodes. Figure 2. 19,a shows a gate for four input signals. An output signal is produced only when all the four input signals are applied. When the impedances of the signal sources are sufficiently low, the circuit can be simplified and takes the form shown in Figure 2.19,b.

c) OR gate circuits

A triode OR circuit is shown in Figure 2. 20. Here an output signal is produced when a signal is received at either inputs. A transistor OR circuit is shown in Figure 2. 21. The simplest OR gate uses diodes. For example, Figure 2. 22 shows a three-input OR gate.

Besides the logic elements described up to now, cybernetic systems make wide use of **trigger** circuits.

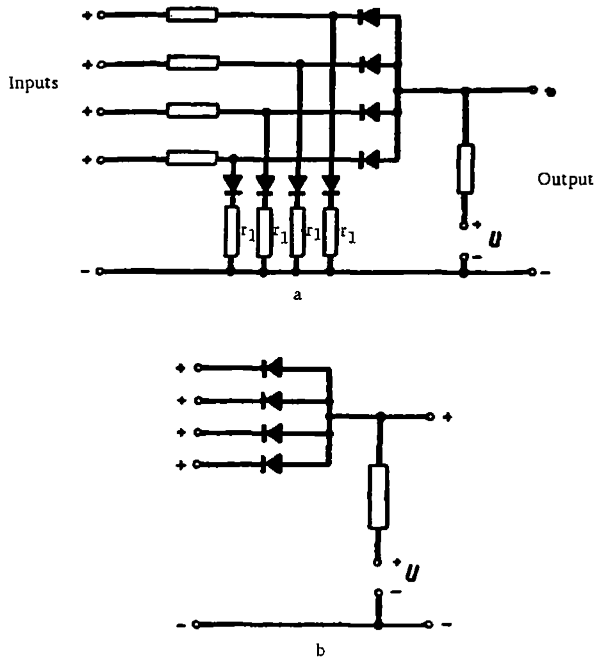

FIGURE 2. 19. Diode AND gates.

One type of trigger circuit, the bistable multivibrator (or flip-flop), has two stable states. One state is denoted by the symbol "0", and the other by the symbol "1". This trigger can therefore be used as a static switch or as a memory cell.

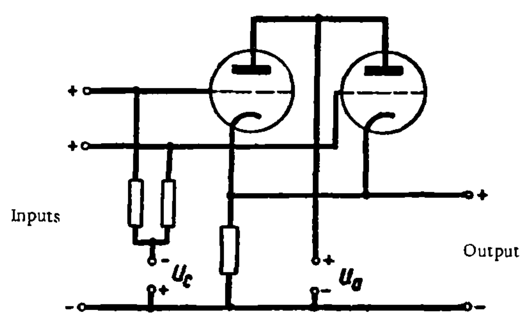

FIGURE 2. 20. Triode OR gate circuit.

A triode bistable multivibrator circuit is shown in Figure 2.23. In this circuit, the grid of each tube is coupled through a resistor to the plate of the other tube.

When the supply voltage u_a is applied to the circuit, the current passing through the two tubes will not be equal because of some difference in their characteristics. Let the higher current pass through T_1. Owing to the greater voltage drop across R_3, the voltage at point a (relative to u_a) will be less than the voltage at point b. The grid voltage of T_2 will therefore be lower than the grid voltage of T_1, which will further decrease the current

through T_2. The voltage at point b will consequently further increase with respect to the voltage at point a. This action continues until T_1 is fully conducting, and T_2 is cut off.

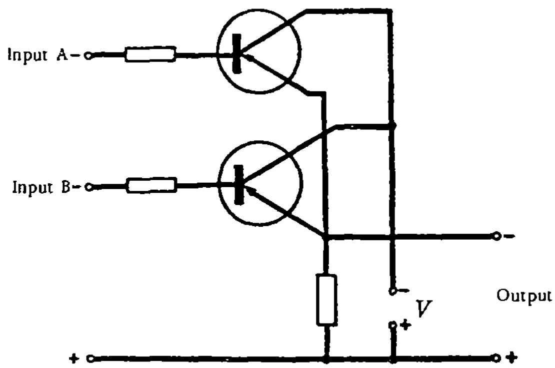

FIGURE 2.21. Transistor OR gate circuit.

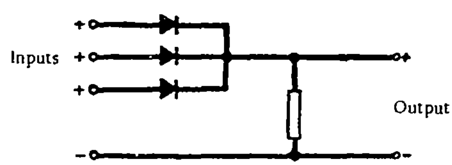

FIGURE 2.22. Diode OR gate.

Since T_1 is fully conducting, the voltage at point a is zero (for the sake of simplicity, the resistance of the tube is neglected). The grid voltage of T_2 will therefore be negative, thus maintaining cut-off. Since T_2 is cut off, the voltage at point b is approximately equal to $+u_a$. The voltages and the resistors in the circuit are so chosen that the grid voltage of T_1 is positive and the tube remains fully conducting.

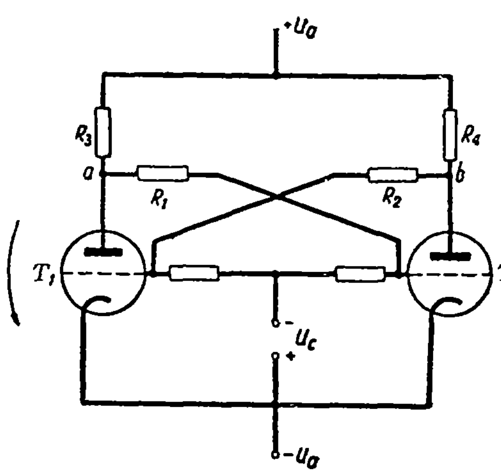

FIGURE 2.23. Bistable multivibrator circuit.

Both tubes will remain in their respective states until further triggering action takes place.

A bistable multivibrator can also be assembled on transistors. One of these circuits is shown in Figure 2.24.

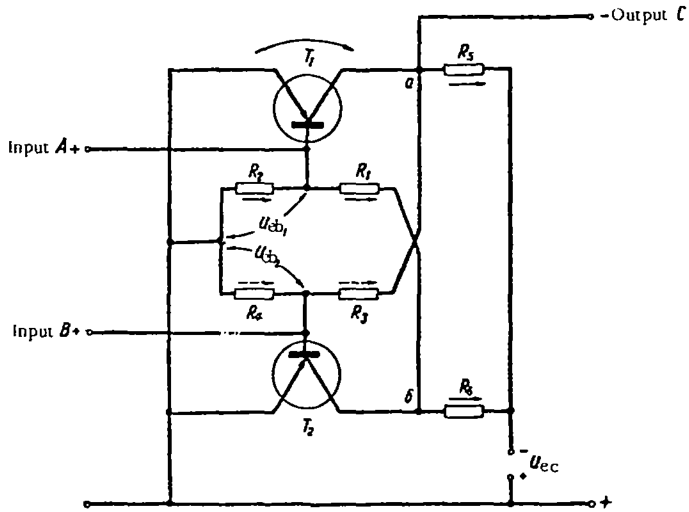

FIGURE 2.24. A transistor bistable multivibrator.

Let transistor T_1 be saturated and transistor T_2 be cut off. Neglecting the emitter-collector resistance, the voltage at point a may be taken as equal to $+u_{ec}$. In this case, no current flows through R_3 and R_4; the bias voltage of T_2 (u_{eb_2}) is zero and T_2 is cut off. Since T_2 is cut off, the voltage at point b is equal to $-u_{ec}$. A current i_1 from the supply u_{ec} flows through R_2, R_1, and R_6 (solid arrows in Figure 2.24). The bias voltage of T_1 is therefore equal to $u_{eb_1} = i_1 R_2$. Resistors R_2, R_1, and R_6 are so chosen that at this voltage T_1 is full on.

Transistor T_1 will remain full on and T_2 cut off until a triggering voltage, equal to, or greater than, u_{eb_1}, is applied to input A. It can be seen from the circuit in Figure 2.24, that in this case, T_1 will be turned off, causing point a to acquire a negative voltage. A current i_2 (dashed arrows) flows through R_4, R_3, R_5, and a positive bias voltage $u_{eb_2} = i_2 R_4$ is now applied to T_2 which is consequently triggered.

If now the voltage applied to input A is disconnected, the transistors remain in their state: T_1 is cut off and T_2 is full on. This is so because the voltage at point b is now $+u_{ec}$ Hence, the bias voltage of T_1 (u_{eb_1}) is zero and the transistor remains cut off. The voltage at point a is maintained at $-u_{ec}$, and the current flow through R_4 maintains the voltage u_{eb_2} keeping T_2 fully on. If now a triggering voltage is applied to input B of the circuit, the circuit is restored to its previous state, i.e., T_1 is turned on and T_2 is cut off.

Thus, when an external signal is received at input A, the trigger circuit passes from one stable state to another and a 1 is produced at output C. This 1 is "cleared" when a signal is received at input B. Consequently, the bistable multivibrator is widely used in automatic control systems as a memory cell and as a static relay (switching element).

It should be noted that a bistable multivibrator is not an independent logical element. It can be designed using OR and NOT gates. Consider the circuit shown in Figure 2.25. Assume that the output *C* is 1, and no signals are received at inputs *A* and *B*. The 1 from output *C* is fed back to the input of the OR gate. Since a 1 is now received at the input of OR_1, the input of NOT_2 is 0. The output from the circuit will be set to 1 indefinitely. In other words, the circuit "stores" a 1.

Let now a signal be fed to input *B*. The input of NOT_2 is now also 1. The circuit output (*C*) is thus cleared to 0. In this case, signal *B* "cleared" the information stored in the memory cell.

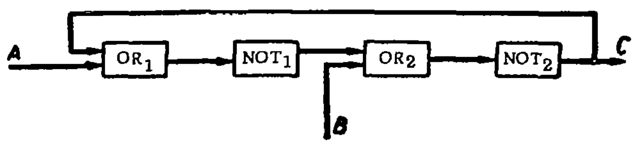

FIGURE 2.25. Bistable multivibrator designed of OR and NOT gates.

When the signal *B* is switched off, the output remains at 0. This 0 is transmitted by the feedback loop to the input of the OR_1 gate. Now the input of the NOT_2 gate is 1. Hence, after the signal at *B* has been lifted, the circuit remains cleared to 0. When a signal is received at input *A*, the output *C* is again set to 1. When signal *A* is lifted, the circuit remains set to 1. In this case we say that the circuit "stores" a one.

A memory cell is represented symbolically by a rectangle with a letter *M* inscribed (Figure 2.26). When a signal is received at input *A*, the circuit stores the signal. When a signal is received at input *B*, the output is cleared to 0.

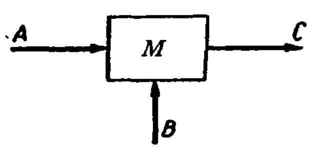

FIGURE 2.26. Symbolic representation of a memory cell.

Besides bistable multivibrators, memory cells may also be made of electrostatic, magnetic, and electromechanical elements.

4. DIODE MATRICES

A **diode matrix** is a network of horizontal and vertical conductors, called wires (Figure 2.27), connected together by semiconductor diodes at predetermined intersection points. The horizontal wires are generally used as inputs, and the vertical as outputs. For example, if we short-circuit the diodes in Figure 2.27 and apply a positive voltage to input *a*, a signal is produced at every output. If the short-circuits are removed, a signal will be produced at the third and the first outputs only.

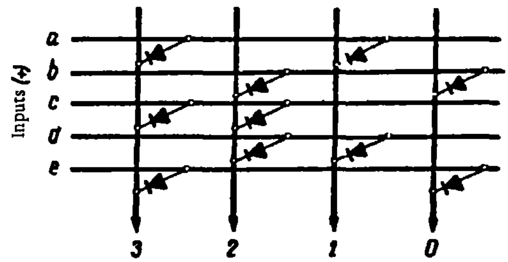

FIGURE 2.27. Diagram of a diode matrix.

Since a diode matrix "stores" the rules (a program) for the gating of input signals, it is often called a diode-matrix memory. Depending on the connections, the diode matrix will perform various logical operations in automatic-control systems. Let us now consider the most common applications of the diode matrix.

a) Logical OR circuit

Each horizontal wire of the matrix (Figure 2.27) is the input to an OR gate. For example, output 3 performs the logical operation

$$a \text{ OR } c \text{ OR } e.$$

This matrix can be used as a binary encoder. Let each vertical wire represent a certain digit of a binary number. When a positive voltage is applied to input a, signals are produced at outputs 1 and 3, and represent the binary number 1010. Input a is thus assigned the number 1010, or in other words input a is coded as 1010. The other inputs are similarly coded. The code for the signals received at the inputs is given in Table 2.2.

TABLE 2.2
Coding of signals

Input	a	b	c	d	e
Code	1010	0101	1100	0110	1001

Consider the use of an OR matrix for automatic-signal control in a railway terminal, with the timetable as given in Table 2.3.

The automatic-signal control is realized as follows. A timer is connected to the matrix input (Figure 2.28). This timer applies a positive voltage to input a, at 12.37 hrs, to input b at 2.05 hrs, etc. When this voltage is applied, the diode matrix produces a binary-coded signal. The decoder interprets this signal and switches-in the appropriate lights.

TABLE 2.3

Train timetable

Departure time	Destination	Type	Platform No.
12.37	Riga-Moscow	Express	1
14.05	Riga-Leningrad	Express	3
16.38	Riga-Simferopol'	Local	2
17.41	Riga-Liepaya	Local	1

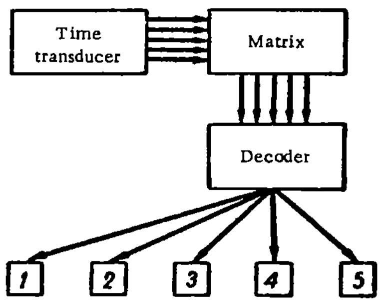

FIGURE 2.28. Automatic-signal control.

The train destinations can be coded as follows:

Riga-Moscow	001
Riga-Leningrad	010
Riga-Simferopol'	011
Riga-Liepaya	100

The type of train and the number of platform of departure are coded similarly:

Express	01	Platform No. 1	01
Local	10	Platform No. 2	10
		Platform No. 3	11

Since the information as to the destination, the type of train, and the platform number must be available simultaneously, all the three matrices may use common inputs.

The resulting matrices are shown in Figure 2.29. At 12.37 the timer applies a positive voltage (1) to input *a* of this matrix. The first matrix (destination) thus produces an output signal 001, the code of the Riga-Moscow train. The second matrix (train type) produces an output signal 01, which corresponds to the information "express". Finally, the third matrix specifies platform No. 1 (01). The matrices function similarly when signals are fed to the other inputs.

If the information on the trains' movements changes, the matrix array is correspondingly modified. For example, if the Riga-Moscow train must be

dispatched from the second platform and not the first, the diode connecting input *a* with output 0 in the "platform number" matrix must be removed, and used to connect input *a* with output 1.

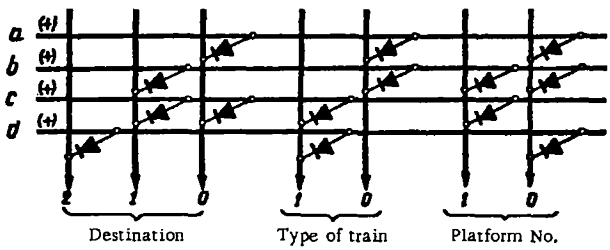

FIGURE 2. 29. Encoding matrix.

If there is no need to modify the matrix array, the diodes are soldered in position. If, however, the array must be often modified, the diodes are connected by means of screws and fasteners.

The binary code produced at the matrix output must be decoded. A simple electromagnetic relay circuit for decoding the platform is shown in Figure 2.30. The matrix producing the platform-number information has two outputs (representing two-bit numbers). The decoder therefore requires two relays. The voltage from the output corresponding to the first-order bit is applied to the coil of the left relay, and the voltage from the output corresponding to the zero-order bit is applied to the right relay. When no voltage is applied to the coil, the contacts remain in the upper position, which represents 0. When voltage is applied to the coil, the relay contacts switch over to the lower position which represents 1.

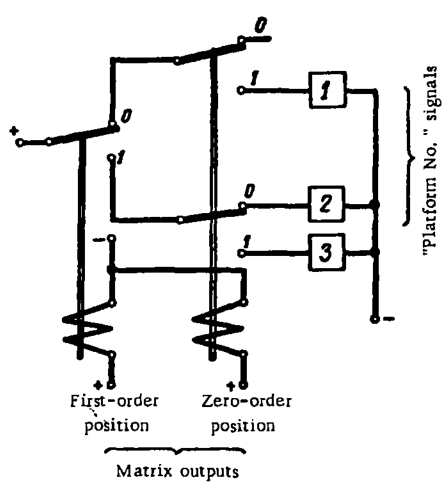

FIGURE 2. 30. A decoder for the information on platform numbers.

The decoder functions as follows. Let a signal 10 be produced by the "platform number" matrix. The left relay is energized. while the right remains de-energized, Light appears under "Platform No. 2". The decoder functions similarly when other codes are received from the matrix.

Decoders for the "destination" and the "train type" matrices are designed similarly to the decoder of the "platform number" matrix.

In some cases the number of matrix outputs are equal to the number of facilities to be switched by the matrix. In this case no decoder is needed and the output signals are fed directly to the facilities controlled. An example of such a matrix is a programmed device for the automatic control of parallel transformers.

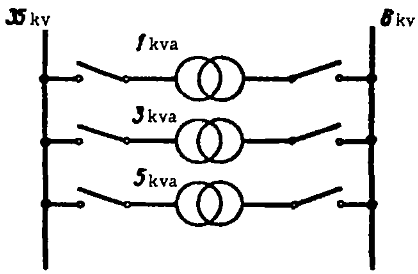

FIGURE 2. 31. Schematic diagram of a transformer substation.

A transformer substation has three transformers of 1, 3, and 5 kva for stepping down 35 kv to 6 kv (Figure 2. 31). The problem is to decide which transformers are to be switched in for a given power flow through the substation. This would save electric power by lowering the no-load losses in the transformers. For example, if 3 kva is channeled through the substation, only the 3 kva transformer should be switched in, while the other transformers should remain disconnected.

A commutator is designed which senses the power flow through the substation, and actuates the corresponding input contact (Figure 2. 32). If the power is 1 kva, contact 1 is closed; if the power is 2 kva, contact 2 is closed, etc. Since the maximum power through the substation is 9 kva, there is a total of nine such contacts.

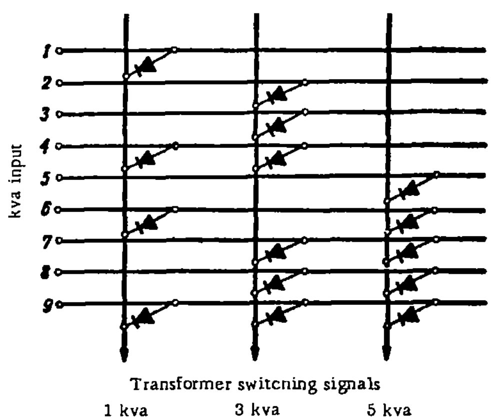

FIGURE 2. 32. A matrix for transformer control.

The contacts of this commutator are connected to the nine inputs (Figure 2.32). The vertical wires of the matrix are connected to the switches controlling the transformer connections to the bus bars of the substation. The diodes in the matrix are so arranged that the power of the actuated transformers is equal to the power channeled through the station. For example, if the power demand is 6 kva, input number 6 is connected by the appropriate diodes to the first and third vertical wires, which switch-in the 1 kva and 5 kva transformers.

The matrix in Figure 2.32 performs the following logical operations:

a) connects the 1 kva transformer if the power channel through the substation is 1, OR 4, OR 6, OR 9 kva;

b) connects the 3 kva transformer if the power is 2, OR 3, OR 4, OR 7, OR 8, OR 9 kva;

c) connects the 5 kva transformer if the power is 5, OR 6, OR 7, OR 8, or 9 kva.

If the inputs and the outputs in the previous matrices are interchanged, we obtain matrices performing the logical AND operation.

b) Logical AND circuit

A matrix performing the logical AND operation is shown in Figure 2.33. This matrix, unlike the matrix shown in Figure 2.27, converts binary information into a control signal. It is therefore called a d e c o d e r.

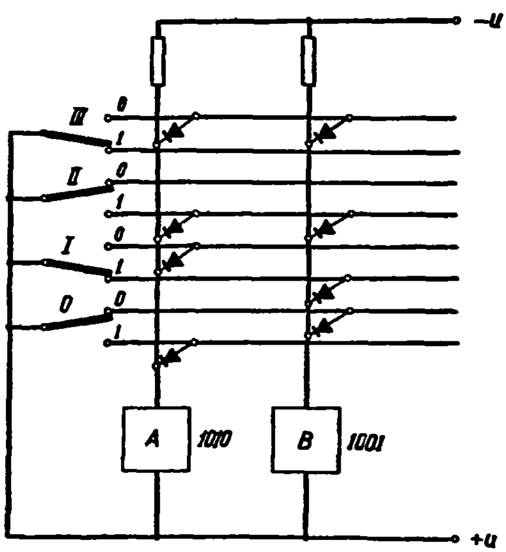

FIGURE 2.33. A matrix performing the logical AND operation.

Four-bit information is fed into the relays of contacts 0, I, II, and III, which are the matrix inputs. For example, suppose that device *A* must be energized when the binary signal 1010 is received. Device *A* is connected

to a vertical wire which is connected by diodes to the appropriate inputs, as shown in Figure 2.33. It receives a signal only if contacts III and I are in the lower position, and contacts II and 0 are in the upper position. Otherwise, device ***A*** is shunted by one of the diodes and the voltage applied to it is zero. Device ***B*** is similarly controlled by the signal 1001.

The number of vertical wires in the decoder is determined by the number of devices to be controlled. The decoder may therefore have one or more vertical wires. The number of horizontal wires required is twice the number of bits in the binary-coded number.

The diodes are connected in the matrix (Figure 2.33) according to the following rule: the diodes connected to the vertical conductor must gate a number which is the negation of the input code. For example, if device ***A*** is triggered when the number 1010 is fed into the matrix, the diodes connected to its vertical wire must gate the number 0101.

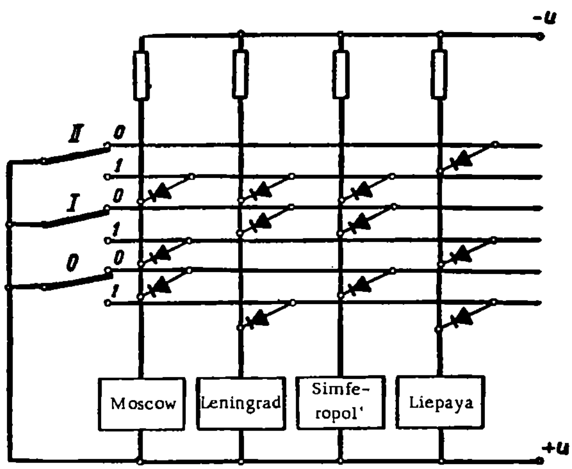

FIGURE 2.34. Decoder for "destination" signals.

Let us consider a matrix performing the AND operator and decoding the binary signal "destination" (Figure 2.29).

As we have previously shown, the matrix in Figure 2.29 generates the following codes:

Riga-Moscow	001
Riga-Leningrad	010
Riga-Simferopol'	011
Riga-Liepaya	100

Since a three-bit code is used to control four different signal lights, the matrix must have three double inputs and four outputs, as shown in Figure 2.34.

5. DESIGN OF ENCODING MATRICES

The numbers encoded by a diode matrix can be written in normal sequence, namely 1, 10, 11, 100, etc. However, to minimize the number of diodes used in the matrix, numbers containing the greatest number of ones are often omitted. In the first case we have a *normally filled matrix*, and in the second case a *matrix with number gaps*.

As we have previously shown in (1.1), the amount of numbers which

a) Calculation of the total number of diodes required in a matrix

We should first observe that an R-bit number is always written using R binary digits with a 1 in the most significant position. For example, three-bit numbers $(R=3)$ are written as follows:

$$\begin{array}{c} 100 \\ 101 \\ 110 \\ 111 \end{array} \tag{2.14}$$

Such numbers as 010 and 011, although written using three bits, are in effect two-bit numbers. Leading zeros are written to facilitate comparison with three-bit numbers. Still, it should be kept in mind that in R positions we may represent not only R-bit numbers, but also all the numbers having less than R positions. For example, in three positions we may also write the two- and one-bit numbers:

$$\begin{array}{cc} 010 & 000 \\ 011 & 001 \end{array}$$

As we have previously shown in (1.1), the amount of numbers which can be represented by R digits is given by the series

$$N=2^0+2^1+2^2+\ldots+2^{R-1}=\sum_{k=0}^{R-1} 2^k.$$

From this expression it can be seen that in the R-th position we have the following amount of R-bit numbers:

$$n_R=2^{R-1}. \tag{2.15}$$

Example. How many three-bit numbers $(R=3)$ are there?

$$n_3=2^{3-1}=4.$$

Consider the structure of the R-bit numbers in (2.14). All these numbers have in their most significant position a 1, each one of which requires a diode. In all the other positions, the number 1 (which requires a diode) occurs in only half the number of digits. Therefore the number of diodes required to code these R-bit numbers may be expressed by the equation

$$D_R=2^{R-1}+\frac{1}{2}\,[2^{R-1}(R-1)].$$

Simplifying further, we derive the expression

$$D_R=2^{R-2}(R+1). \tag{2.16}$$

Since all the numbers of a lower order than R can also be represented in R positions, the total number of diodes in an R-bit matrix is given by the equation

$$D=\sum_{m=1}^{R} 2^{m-2}(m+1). \tag{2.17}$$

Example. Determine the number of diodes required for a four-bit matrix $(R=4)$.

$$D=\sum_{m=1}^{4} 2^{m-2}\,(m+1)=2^{-1}\cdot 2+2^{0}\cdot 3+2^{1}\cdot 4+2^{2}\cdot 5=32.$$

b) Calculation of the amount of R-bit numbers having the same number of diodes

In an R-bit matrix we have numbers represented by 1, 2, 3, ... R diodes. If the number of diodes is denoted by the letter x, the following inequality is always satisfied:

$$R\geqslant x\geqslant 1. \tag{2.18}$$

Let us develop a procedure for determining how many R-bit numbers have the same number of diodes. Take for example the case where R = 4, where there is a total of eight four-bit numbers:

1000
1001
1010
1011
1100
1101
1110
1111

These numbers are divided into four groups and written in a different sequence:

1	000	I
1	001	
1	010	II
1	100	
1	011	
1	101	III
1	110	
1	111	IV

An examination of these groups shows that among R-bit numbers there is but a single number using R diodes, and a single number using one diode:

$$n_R^{1D}=1; \tag{2.19}$$

$$n_R^{RD}=1. \tag{2.20}$$

The second and the third group, omitting the 1 in the most significant positions, are in fact permutations of one or two diodes in three positions. Generalizing, we may say that these numbers are obtained by permutating from one to R-2 diodes in R-1 positions.

The theorem of combinations known from elementary algebra states that the number of all possible combinations of n elements taken m at a time is given by the equation

$$C_n^m=\frac{n!}{m!(n-m)!}.$$

The amount of numbers having x diodes is therefore equal to

$$n_R^{xD}=\frac{(R-1)!}{(x-1)!(R-x)!};\ (R>x>1). \tag{2.21}$$

The expression $(R>x>1)$ indicates that this equation applies for all x less than R and greater than 1. The difference $(x-1)$ enters the denominator of this equation, since the 1 diode occupying the most significant position is not permutated. Since this diode is being disregarded, the most significant position must also be disregarded in calculating the permutation. Therefore, the number of positions in the numerator of (2.21) is also reduced by one, i.e., R-1.

Example. Determine the amount of five-bit (R=5) numbers having 1, 2, 3, 4, and 5 diodes.

The number of numerals having 1 and 5 diodes is determined by (2.19)

and (2.20), and the number of numerals having 2, 3, and 4 diodes are calculated using (2.21)*:

$$n_5^{xD}=\frac{(R-1)!}{(x-1)!(R-x)!}=\frac{4!}{(x-1)!(5-x)!}.$$

We thus have

x	1	2	3	4	5
n_5^{xD}	1	4	6	4	1

c) Calculation of the amount of numbers having x diodes in R positions

We have already observed that R positions will represent not only R-bit numbers, but also all numbers of lower order than R.

In (2.20) we noted that R diodes occur in a single number only, that consisting of R ones:

$$N_R^{RD}=1. \tag{2.22}$$

R-bit numbers (see (2.19)), (R - 1)-bit numbers, (R-2)-bit numbers, etc., all have a single number with one diode. Therefore in R positions we have the following number of numerals with one diode:

$$N_R^{1D}=R. \tag{2.23}$$

Let us now consider the cases when $R>x>1$.

The amount of R-bit numbers having x diodes is determined by (2.21):

$$n_R^{xD}=\frac{(R-1)!}{(x-1)!(R-x)!}.$$

Similarly, the number of [(R - 1) - 1]- bit numerals having x diodes is

$$n_{R-1}^{xD}=\frac{(R-2)!}{(x-1)!(R-x-1)!}.$$

* [Actually (2.21) is valid for $x=1$, as $0!=1$. Alternatively, if (2.21) is written

$$C_n^m=n_R^{xD}=\frac{R!}{x!(R-x)!}\quad (R\geq x>1)$$

it will give the amount of numbers having x diodes for all values of x except 1.]

The same applies for all lower-order numbers,

$$n_{R-2}^{xD}=\frac{(R-3)!}{(x-1)!(R-x-2)!};$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$n_{x+1}^{xD}=\frac{x!}{(x-1)!\,1!}.$$

Therefore the total number of numerals having x diodes is given by the equation

$$N_R^{xD}=1+\sum_{m=x+1}^{R} n_m^{xD}. \tag{2.24}$$

The term Σn determines the number of numerals obtained by the permutation of all the numbers from $x+1$ to R. The 1 is added to this term to include the one additional number having all ones and has x diodes. It should be noted that the lowest order still having numbers with x diodes can be determined by the expression

$$R_{\min}=x+1.$$

Inserting for n_m^{xD} in (2.24), we obtain

$$N_R^{xD}=1+\frac{1}{(x-1)!}\sum_{m=x+1}^{R}\frac{(m-1)!}{(m-x)!};\ (R>x>1). \tag{2.25}$$

Let us consider three particular cases of this equation. Substitute $x=R-1$, $R-2$, and $R-3$ in (2.25). After simple manipulations we obtain

$$N_R^{(R-1)D}=R; \tag{2.26}$$

$$N_R^{(R-2)D}=\frac{R(R-1)}{2}; \tag{2.27}$$

$$N_R^{(R-3)D}=1+\frac{(R-3)(R^2+2)}{6}. \tag{2.28}$$

All lower-order expressions can be derived similarly.

Example. Determine how many numbers represented in seven positions ($R=7$) have 6, 5, and 4 diodes.

From (2.26), we see that six $(R-1=6)$ diodes occur in seven numbers. Inserting $R=7$ into (2.27) and (2.28), we find that five diodes occur in twenty-one numbers, and four diodes in thirty-five numbers.

d) Matrices with one extra position

It is sometimes desirable to introduce an extra position into the matrix, and thus avoid the necessity of representing numbers with numerous diodes. A certain economy of diodes is thus achieved.

As we have previously seen in (1.3), the number of positions required for the sequential representation of N numbers is equal to

$$R=\log_2 (N+1).$$

To simplify the presentation, we shall assume that R is an integer. We introduce an additional position into the matrix

$$R_1=R+1.$$

The addition of one position increases the number of numerals that can be represented in the matrix (see (1.1)) by

$$\Delta N=2^{R_1-1}. \qquad (2.29)$$

We can thus remove from the matrix ΔN numerals requiring the largest number of diodes.

The maximum number of diodes R occurs (see (2.22)) in a single numeral only. R-1 diodes occur in $N_R^{(R-1)D}$ numerals, R-2 diodes in $N_R^{(R-2)D}$ numerals, etc. Since the total number of numerals which can be removed (see (2.29)) is 2^{R_1-1}, we have

$$2^{R_1-1}=1+N_{R_1}^{(R_1-1)D}+N_{R_1}^{(R_1-2)D}+\ldots+N_{R_1}^{(R_1-k+1)D}+qN_{R_1}^{(R_1-k)D}$$

or

$$2^{R_1-1}=1+\sum_{m=1}^{k-1} N_{R_1}^{(R_1-m)D}+qN_{R_1}^{(R_1-k)D}. \qquad (2.30)$$

In addition, this equation shows the order in which the 2^{R_1-1} numbers are eliminated. First the single number with R_1 diodes is removed. Then all numbers with R_1-1 diodes ($N_{R_1}^{(R_1-1)D}$) are eliminated, followed by all numbers with R_1-2 diodes, etc. However, the last group of numbers, those having R_1-k diodes, cannot always be removed in its entirety. We therefore remove only part (q) of these numbers. The coefficients k and q are so chosen that the left-hand side of equation (2.30) is equal to the right-hand side.

From (2.16), we see that the introduction of an additional position into a matrix in which all the numbers are represented sequentially, increases the number of diodes required.

$$\Delta D_1=2^{R_1-2}(R_1+1). \qquad (2.31)$$

On the other hand, from (2.30) we see that the omission of 2^{R_1-1} numerals having the highest number of diodes enables us to save ΔD_2 diodes:

$$\Delta D_2=R_1+\sum_{m=1}^{k-1}(R_1-m)N_{R_1}^{(R_1-m)D}+q(R_1-k)N_{R_1}^{(R_1-k)D}. \qquad (2.32)$$

Thus, the net saving in diodes is equal to

$$\Delta D=\Delta D_2-\Delta D_1=R_1+\sum_{m=1}^{k-1}(R_1-m)N_{R_1}^{(R_1-m)D}+$$

$$+q(R_1-k)N_{R_1}^{(R_1-k)D}-2^{R_1-2}(R_1+1). \qquad (2.33)$$

Example. *From* (1.3) *we have seen that* 31 *numbers in sequence can* be represented in

$$R=\log_2(31+1)=5 \text{ positions}.$$

Find the number of diodes saved if one additional position is introduced into the matrix $(R_1=R+1=6)$.

1. *Find the coefficients* k and q :

a) the left-hand side of equation (2.30) is equal to

$$2^{R_1-1}=2^5=32;$$

b) the terms in the right-hand side of (2.30) are obtained from (2.26), (2.27), and (2.28);

$$N_{R_1}^{(R_1-1)D}=R_1=6;$$

$$N_{R_1}^{(R_1-2)D}=\frac{R_1(R_1-1)}{2}=15;$$

$$N_{R_1}^{(R_1-3)D}=1+\frac{(R_1-3)\ (R_1^2+2)}{6}=20;$$

c) equating the left- and the right-hand sides of (2.30); inserting the above values:

$$32=1+6+15+q\cdot 20.$$

Hence, $k=3$, $q=0.5$.

2. The saving achieved (see (2.22)):

$$\Delta D=R_1+\sum_{m=1}^{k-1}(R_1-m)N_{R_1}^{(R_1-m)D}+$$

$$+q(R_1-k)N_{R_1}^{(R_1-k)D}-2^{R_1-2}(R_1+1)=$$

$$=6+5\cdot 6+4\cdot 15+0.5\cdot 3\cdot 20-16(6+1)=$$

$$=126-112=14 \text{ diodes}.$$

The total number of diodes in a R-equation matrix (2.17) is

$$D=\sum_{m=1}^{R}2^{m-2}(m+1)=2^{-1}\cdot 2+2^0\cdot 3+2^1\cdot 4+2^2\cdot 5+2^3\cdot 6=80 \text{ diodes}.$$

We have thus reduced the number of diodes by

$$\frac{\Delta D}{D}\cdot 100=\frac{14\cdot 100}{80}=17.5\%\text{ diodes.}$$

6. STATIC SWITCHING DEVICES IN DIGITAL AUTOMATIC-CONTROL SYSTEMS

Until recently, most switching operations in automatic-control systems were carried out by electromagnetic relays which involved moving mechanical contacts. These contacts considerably lowered the reliability of automatic-control systems, and the inductance of the relay coils increased the response time of the systems.

The development of electronic logical elements and static switching devices now makes it possible to replace the cumbersome relay mechanisms by more reliable, lightweight, compact, and high-speed elements. Some of the static elements used in automatic-control circuits are vacuum and gas-discharge tubes, magnetic amplifiers, and semiconductor devices. Although magnetic amplifiers and semiconductor devices are not ideal circuit breakers (the current does not drop to zero when a circuit is disconnected by these elements), they are very popular because of their ruggedness and long service life.

The use of standard logical elements in automatic-control systems considerably lowers the time required for system design. In addition, their use facilitates maintenance since a new standard element can simply be plugged-in to replace a malfunctioning component.

A recent advance in the design of electronic logical elements is the use of *solid circuits*. In these circuits there are no individual resistors, diodes, transistors, etc. A solid circuit is generally a germanium or silicon plate whose area is no greater than the cross section of a match head. Diodes, transistors, capacitors, and resistors are formed on this plate using etching, diffusion, and other techniques. A single crystal without any connecting wires is thus capable of performing very complicated logical transformations. Since solid circuits are not soldered, they are exceptionally reliable and their use is expected to become very widespread in automatic-control systems.

Let us first consider the simplest relay circuits and their equivalent logical elements. Since these logical elements perform all the functions of the relay circuits, they can be called *logical equivalents*.

Consider the circuit in Figure 2.35 where three relays are connected in series, and whose contacts are normally open. Current will flow in the circuit only when voltage is applied to the coils of all the three relays (A, B, and C). The relays thus perform the following logical operation:

$$D=ABC.$$

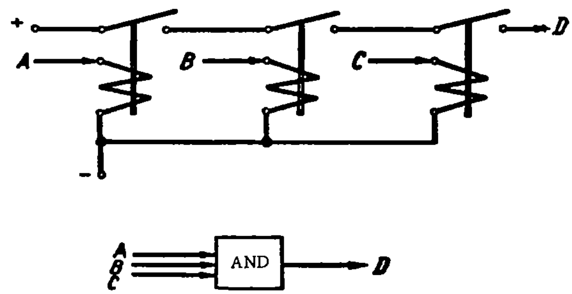

FIGURE 2.35. Logical equivalent of three relays whose contacts are connected in series.

This logical operation can be performed using a single AND element (Figure 2.35)

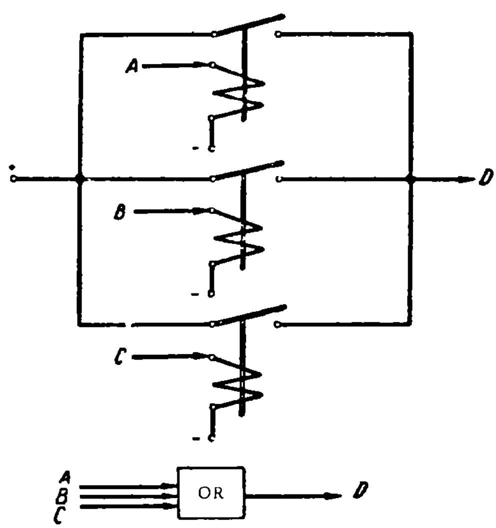

FIGURE 2.36. The logical equivalent of three relays whose contacts are connected in parallel.

If the contacts of the three relays are connected in parallel (Figure 2.36), they perform the logical operation $D=A+B+C$. They can therefore be replaced by an OR element.

Consider now the case of three relays whose normally closed contacts are connected in series (Figure 2.37). Since the equivalent logical operation is

$$D=\overline{A+B+C},$$

it can be performed by using two elements: OR and NOT.

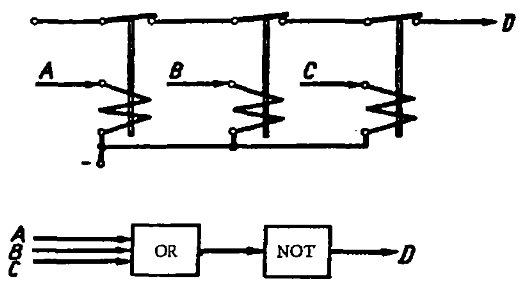

FIGURE 2.37. The logical equivalent of three relays whose *normally closed contacts are connected in series.*

Relays whose normally closed contacts are connected in parallel (Figure 2.38) perform the logical operation

$$D=\overline{ABC}.$$

They can therefore be replaced by a circuit consisting of AND and NOT elements.

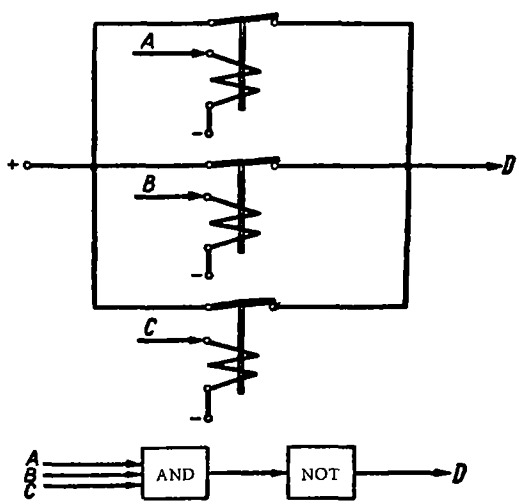

FIGURE 2.38. The logical equivalent of three relays whose normally closed contacts are connected in parallel and their symbolic equivalent.

A switch (Figure 2.39) performs two logical operations:

$$B=\bar{A};\quad C=A.$$

It can therefore be replaced by a NOT element.

Self-holding relays (Figure 2.40) are very frequently used in automatic-control systems. A characteristic feature of these switches is that by means

of a self-holding contact they impress voltage on their own coil, so that no change-over occurs when the signal A is removed. The relay is reset by means of push button K. The self-holding relay performs the following logical operations:

$$B = \overline{C}; \quad C = A + C.$$

These operations (provided the inputs and the outputs of the logic elements are appropriately matched) can be performed by the logic circuit shown in Figure 2.40.

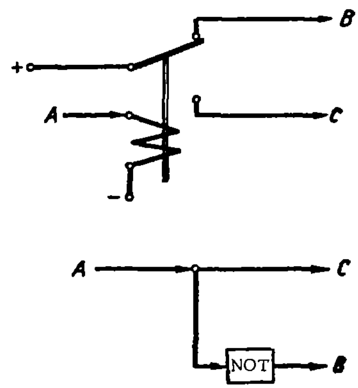

FIGURE 2.39. The logical equivalent of a switch.

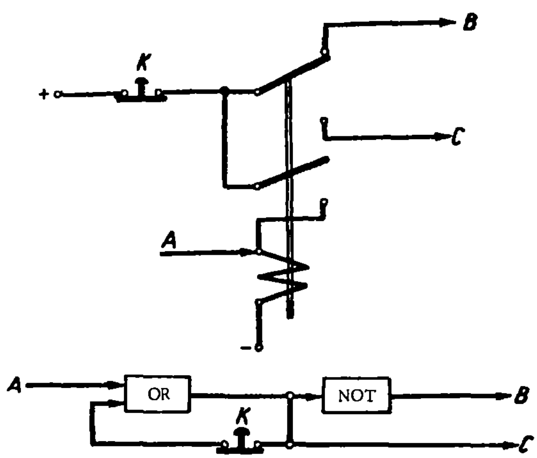

FIGURE 2.40. The logical equivalent of a self-holding relay.

The more complicated the relay circuit, the greater the number of logic elements required in the equivalent circuit. The choice of these elements and their interconnections are made by using Boolean algebra to describe the relay circuit.

At present, universal logic elements are being put into production which will enable the design of any composite logic circuit. These elements are manufactured as printed-circuit units which are very convenient for mounting and adjusting. Versatility is ensured by producing these circuits in units that can be used in many different combinations.

Consider, for an example, the design of a universal AND gate (Figure 2.41). The gate consists (Figure 2.41,a) of three identical sections. A_1 to A_6 are the inputs, B_1 to B_3 are the outputs. In Figure 2.41, b, terminals B are connected with their corresponding terminals K, forming three AND gates each having two inputs. When the terminals are connected as in Figure 2.41, c, we obtain two AND gates, one with two inputs (A_1 and A_2), and one with four inputs (A_3 to A_6). Connecting the terminals according to Figure 2.41, d, we obtain one AND gate with six inputs.

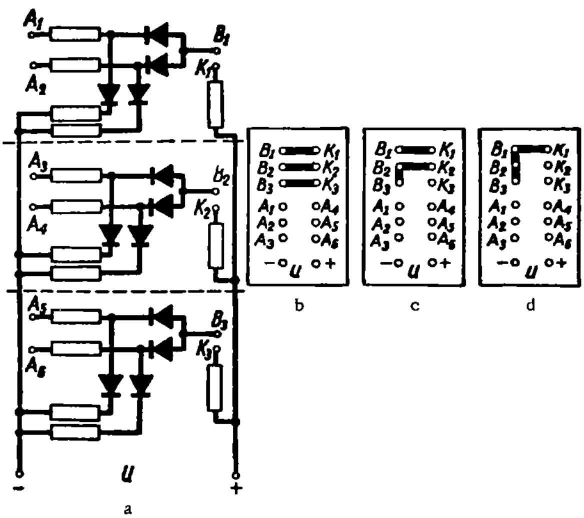

FIGURE 2.41. A universal AND gate: a-internal printed assembly; b, c, d - external terminals and connections.

Using standard NOT, AND, and OR gates cybernetic automatic-control systems may be designed capable of complex logical functions.

To match a logic circuit assembled of universal elements with a large automatic-control system, suitable convertors must be introduced to match the voltages and the currents in the different parts of the system. Let us consider an example of an automatic-control system assembled of logic elements.

Figure 2.42 shows a tank from which water flows continuously. The tank receives water from a pipe controlled by valve K. Our problem is to develop a system which will automatically regulate the level of water in the tank. The valve K should open when the water level drops below point a, and close when the level reaches point b.

Three cases are possible as regards the position of valve K:

a) water level below point a — the valve must be open;

b) water level between points a and b — the valve should either be open

(if the tank is filling) or closed (if the tank is emptying);

c) water level reaching point *b* —the valve must close.

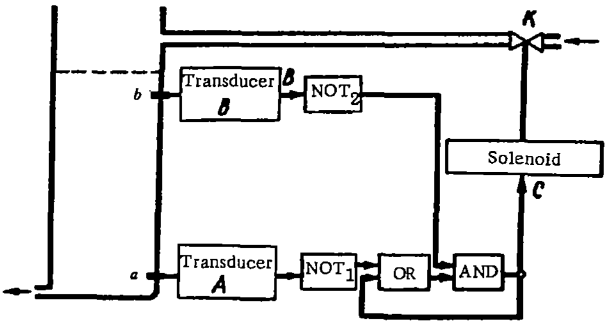

FIGURE 2.42. A system for the automatic regulation of water level in a tank.

Two transducers are used to signal the water level in the tank. The first transducer gives a signal (A) when the water level is at point *a* and higher. The second transducer produces a signal (B) when the water level reaches point *b*.

Then, in accordance with the three cases for the valve, the logical control circuit can be represented as follows:

1) if $A=0$ and $B=0$, the valve should be opened (signal C);

2) if $A=1$ and $B=0$, the position of the valve need not be changed;

3) if $A=1$ and $B=1$, the valve must be closed.

Let us now draw a logic circuit satisfying these conditions. From condition 1 it follows that the signal to open the valve (C) is determined by the expression

$$C=\bar{A}\bar{B}. \tag{2.34}$$

This equation satisfies condition 1. If the water level is below point *a*, $A=0$ and $B=0$. Therefore $C=1$, i.e., a signal is sent to open the valve. However, condition 2 is not satisfied, for when the water level rises above point *a*, the first transducer will emit a signal $A=1$, and according to (2.34) $C=0$. The valve will thus close, although it should remain open. To avoid this situation, (2.34) must be modified as follows: :

$$C=(\bar{A}+C)\bar{B}. \tag{2.35}$$

This logical operation regulates the system as follows:

a) $A=0$, $B=0$; then $C=1$, i.e., when the water level is below point *a*, the valve opens;

b) $A=1$, $B=0$; then $C=1$, i.e., when the level of the water rises above point *a* (after the valve has been opened), the valve remains open; ;

c) $A=1$, $B=1$; then $C=0$, i.e., when water reaches point *b*, the valve closes;

d) $A=1$, $B=0$; then $C=0$, i.e., as the water level drops (after the valve has been closed), the valve remains closed.

The logical proposition (2.35) thus fully satisfies the requirements for automatic regulation of the water level. This proposition (Figure 2.42) is

realized using four gates (two NOT gates, one AND and one OR gate). The input of the AND gate is fed to the solenoid. When voltage is applied to the solenoid ($C=1$), valve K opens; when the signal is lifted ($C=0$), the valve closes. Thus we have demonstrated how to design an automatic-control system by reasoning in terms of logic elements.

7. LOGIC ELEMENTS USING TUNNEL DIODES

Recently, a new semiconductor device called the tunnel diode has been developed. Owing to its simplicity, high-speed operation, and distinctive current-voltage characteristics, the tunnel diode is being increasingly used in amplifiers, oscillators and logic elements.

The static current-voltage characteristic of a tunnel diode (the forward portion) is shown in Figure 2.43. The reverse portion of the characteristic (the dependence of the current passing through the diode on the reverse voltage applied to the diode) is generally not used in logic elements. Unlike ordinary semiconductor diodes, the tunnel diode develops a very small reverse resistance when a reverse voltage is applied.

The forward tunnel-diode characteristic (Figure 2.43) can be divided into three specific regions. The first region extends from the origin to point A. In this range the static resistance of the diode (the ratio of the diode voltage to the diode current) remains fairly constant, increasing slightly near the end of the range (see Figure 2.44). The dynamic resistance of the diode, defined as

$$R_{d}=\frac{du_{d}}{di},$$

is also constant at first (Figure 2.45), but increases sharply to infinity at the end of the first region.

The second region of the tunnel-diode curve (Figure 2.43) is characterized by a decrease in the diode current from i_{max} (point A) to i_{min} (point B). The static resistance of the diode (Figure 2.44) increases in this range, and the dynamic resistance (Figure 2.45) becomes negative.

In the third region of the diode characteristic (Figure 2.43) the current increases sharply with the voltage applied. The static resistance attains its maximum in this range and starts decreasing, whereas the dynamic resistance (Figure 2.45) is again positive.

Figure 2.43 shows that static characteristic of a tunnel diode which is obtained when the voltage is varied very slowly. If the diode voltage is varied more rapidly, we obtain the dynamic current-voltage characteristic. It differs from the static characteristic in that it is affected by the capacitance of the diode. Otherwise, it has approximately the same form as the static curve shown in Figure 2.43.

In the mathematical analysis of circuits using tunnel diodes, the calculations are often simplified by assuming an idealized current—voltage characteristic of the tunnel diode. This characteristic is shown in Figure 2.46.

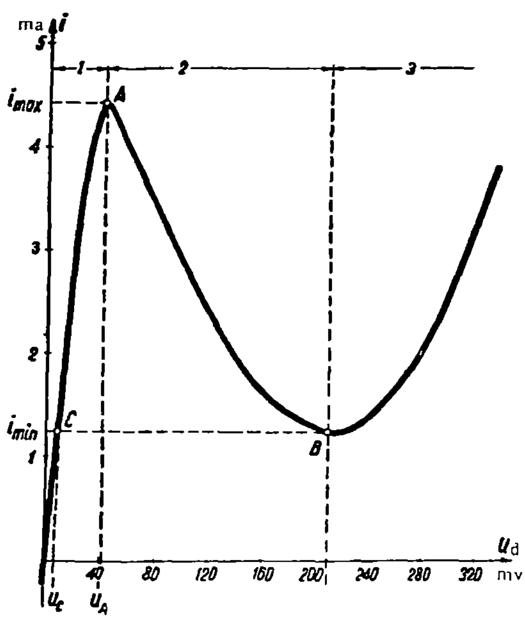

FIGURE 2.43. Tunnel-diode current-voltage characteristic.

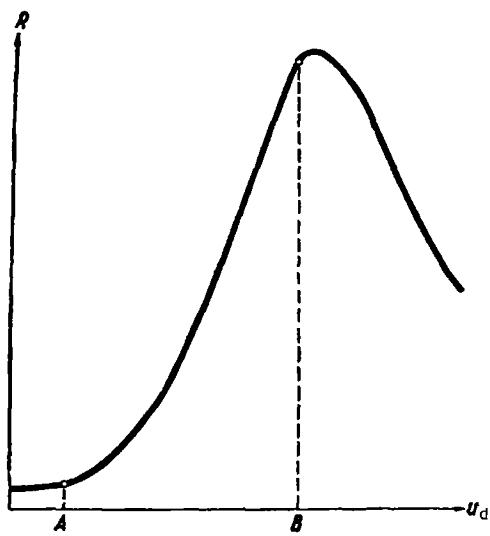

FIGURE 2.44. Static resistance of a tunnel diode.

Symbolically, a tunnel diode is drawn using the common semiconductor diode symbol (Figure 2.47) enclosed in a circle.

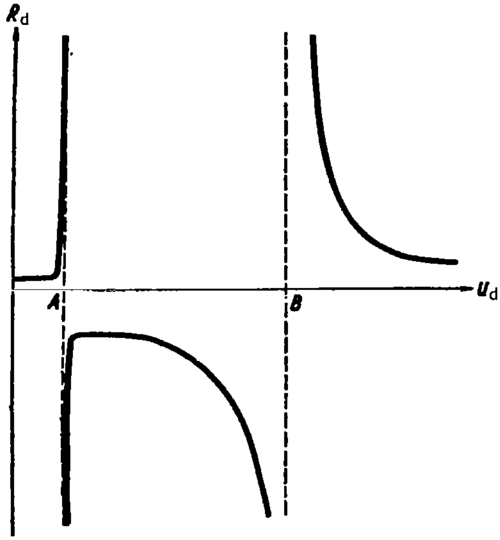

FIGURE 2.45. Dynamic resistance of a tunnel diode.

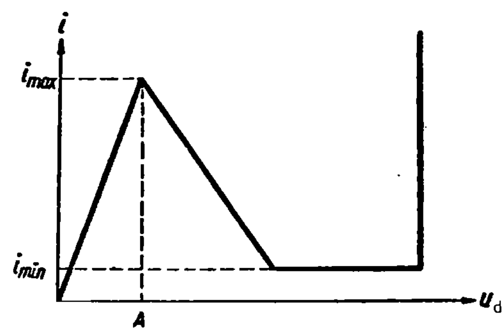

FIGURE 2.46. Idealized tunnel-diode characteristic.

Let us now consider the properties of a simple circuit (Figure 2.48) consisting of a tunnel diode and a resistor connected in series. The same current passes through the diode and the resistor. Therefore, adding the diode voltage (u_d) and the voltage drop across the resistor (u_R) for a given current, we obtain the input voltage of the circuit (u_{bs}). This voltage is generally called the b i a s v o l t a g e.

The current i of the d i o d e—r e s i s t o r circuit plotted as a function of u_{bs} is shown by the dashed curve in Figure 2.49. The solid lines represent the current-voltage characteristics of the diode (u_d) and the resistor (u_R). From the characteristic of the circuit we see that at first the current increases to i_{max} with increasing bias voltage (u_{bs}). Beyond the point A_1 the circuit characteristic reverses itself in the direction of decreasing bias voltage. The bias voltage, however, is independent of the current flowing

in the circuit. Therefore, when voltage increases at point A_1 the current will drop abruptly to the value determined by point B_1. When u_{bs} is further increased, the current again starts increasing. When the bias voltage drops to point B_2, a second abrupt change of current occurs, this time from a lower to a higher value. The portion A_1B_2 is thus unstable, and the current-voltage characteristic of the diode-resistor circuit has the form shown in Figure 2.50. The portions A_1B_1 and B_2A_2 indicate the current jumps which occur at bias voltages of u_{bs_1} and u_{bs_2}, respectively.

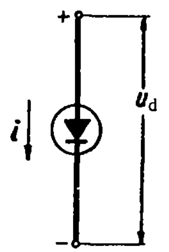

FIGURE 2.47. Symbolic drawing of a tunnel diode.

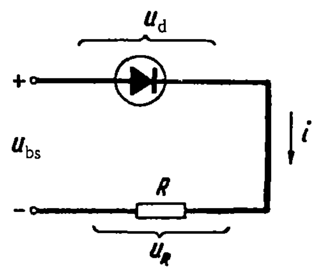

FIGURE 2.48. A tunnel diode in series with a resistor.

In Figure 2.49 we plotted the characteristic of the diode-resistor circuit using the current-voltage characteristics of the diode and the resistor. The reverse problem can be solved similarly: the characteristic of the tunnel diode operating in a diode-resistor circuit can be constructed from the characteristics of the circuit and the resistor. Given several values of current and computing the diode voltage from the bias voltage (Figure 2.50) and the voltage drop across the resistor (Figure 2.49), we construct the curve shown in Figure 2.51:

$$i=f(u_d)=f(u_{bs}-u_R).$$

It can be seen from this curve that when the diode voltage increases at point C_1 the diode parameters experience an abrupt change. Consequently when the diode voltage switches from u_{d_1} to u_{d_2}, the current drops from i_{C_1} to i_{C_2}. When the diode voltage decreases, the switching occurs at point D_2. Here the current increases from i_{D_2} to i_{D_1} and the diode voltage drops from u_{d_3} to u_{d_4}.

Note that the current and voltage parameters of the diode-resistor circuit do not always experience this abrupt change when the bias voltage is varied. When the resistance R is small, the portion A_1B_2 in Figure 2.49 may lose its inflection toward the lower values of u_{bs}. In this case, as u_{bs} increases, the current varies smoothly.

The number and location of the stable points on the current-voltage characteristic of a diode operating (for a given u_{bs}) in a diode-resistor circuit can be determined by two methods.

The first method consists in plotting the current-voltage characteristic of the diode-resistor circuit from the known characteristics of the diode and the resistance (Figure 2.50). The number of stable points of the circuit and the currents at these points are then determined from this characteristic for the given bias voltages (u_{bs}).

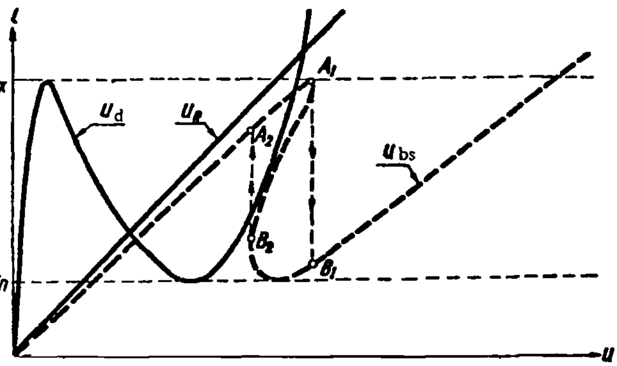

FIGURE 2.49. Construction of the characteristic of the diode-resistor circuit.

For example, if the input voltage of the circuit is u_{bs_K} (Figure 2.50) the circuit has two stable points (K_1 and K_2), with currents i_K and i_{min}, respectively. The current is i_K if the voltage u_{bs} has increased from a lower value to u_{bs_K}. If, however, the voltage u_{bs_K} has been reached as u_{bs} has decreased, the current is i_{min}. When the bias voltage is u_{bs_E}, there is but a single stable point in the circuit, and the current is i_E regardless of the trend of u_{bs} before reaching this point.

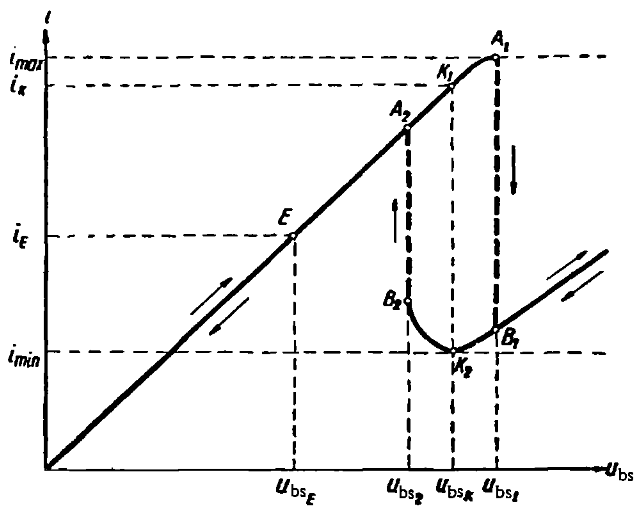

FIGURE 2.50. Working characteristic of the diode-resistor circuit.

Having determined the number of stable points and the currents at these points, we may proceed to determine the stable points of the diode operating in the diode-resistance circuit (Figure 2.51). Since the current in this circuit is equal to the diode current, the stable points of the diode characteristic can be determined from the currents at the stable points of the circuit. For example, if the currents at the stable points in Figure 2.50 are

i_K and i_{min}, then the stable points of the diode in Figure 2.51 are E_1 and E_2.

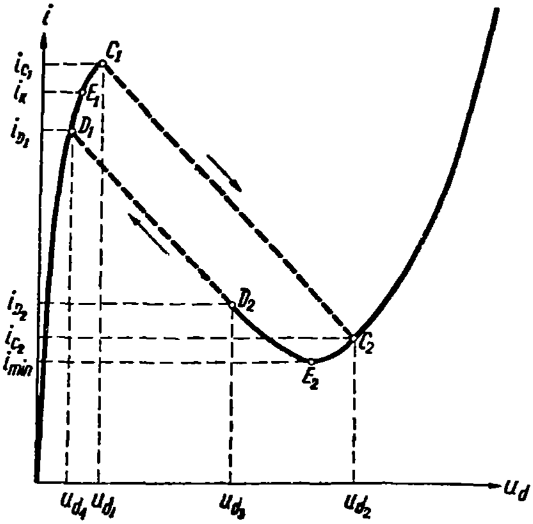

FIGURE 2.51. Current-voltage characteristic of a tunnel diode operating in a diode-resistor circuit.

The second method of determining the stable points on the diode characteristic is simpler. From Figure 2.48, the diode voltage for a given bias voltage (u_{bs}) is given by the equation

$$u_d = u_{bs} - Ri.$$

On the other hand, in Figure 2.43, the diode voltage as a function of the current i is defined graphically:

$$u_d = f(i).$$

The voltage u_d is thus determined by two relationships. The intersection of the straight line $u_d = u_{bs_1} - Ri$ (for a given u_{bs_1}) with the curve $u_d = f(i)$ give the operating points of the diode.

The operating points of the diode can now be determined graphically. We plot (Figure 2.52) the current-voltage characteristic of the diode. The given bias voltage (u_{bs_1}) is laid off the abscissa. A straight line is then passed through the point $i=0$; $u=u_{bs_1}$ at an angle of $\alpha = \operatorname{arcctg} R$. This gives the load line $u_d = u_{bs_1} - Ri$. This line intersects the diode characteristic at points A, B, and C. As previously shown, point B in the second region of the diode characteristic is unstable. Points A and C are stable, and from these points, the corresponding diode currents and voltages can be easily determined.

The diode is switched from one stable point to another in the following manner. Let us assume the diode is operating at point A, and the bias

voltage u_{bs_1} is increased by Δu_1. The load line $u_d = u_{bs_1} + Ri$ is thus displaced parallel to itself to the right and now intersects a single operating point C_1. If now the voltage Δu_1 is removed, both points A and C can be operating points. However, since the diode voltage has decreased, the diode goes to point C. When the bias voltage is decreased by Δu_2, the only operating point is A_1, so that, when the voltage Δu_2 is switched off, the diode operates at point A.

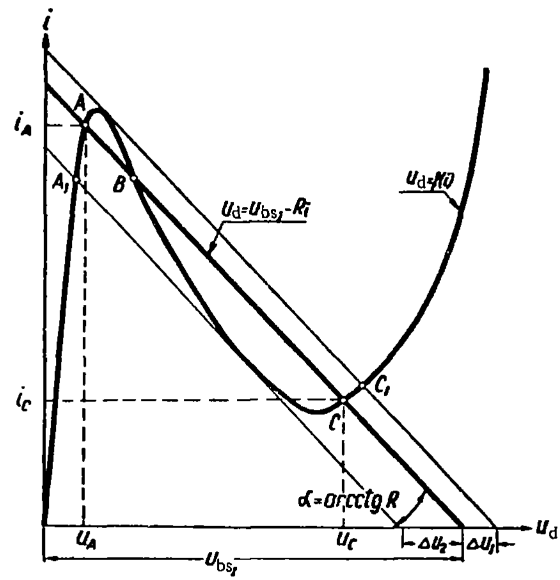

FIGURE 2. 52. The characteristic of a bistable diode—resistor circuit.

When the load in the diode—resistor circuit decreases (from R to R_1), the characteristic of the diode (Figure 2. 53) may have a single stable point only (in this case, point A).

Since the tunnel diode can switch from one state to another, it is suitable for the design of various electronic logic elements. These elements are distinguished by their simple circuitry and high-speed operation (the switching time of tunnel diodes is even shorter than that of vacuum tubes).

Let us now consider how tunnel diodes can be used to design various logic elements. In the circuit shown in Figure 2. 54, the bias voltage (u_{bs_1}) and the resistor R are so chosen that the tunnel diode operates approximately at point A (Figure 2. 52), in the immediate vicinity of maximum current. A still better choice of u_{bs_1} and R would ensure this maximum current. At this point, however, any slight increase of voltage can lead to maloperation of the system. The resistors R_1 and R_2 are so chosen that the voltage across R_2 is equal to the diode voltage u_A. In this case, the output voltage of the circuit is 0.

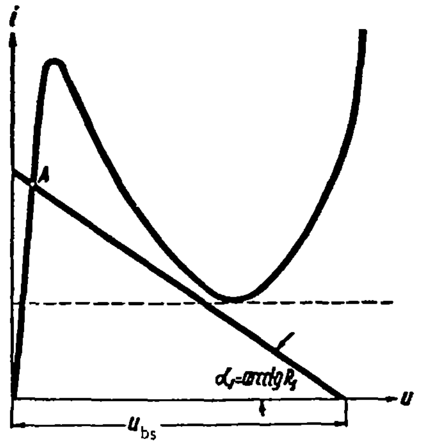

FIGURE 2.53. Monostable characteristics.

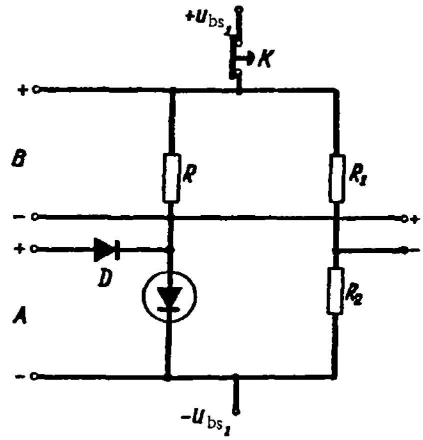

FIGURE 2.54. Tunnel-diode switching circuit.

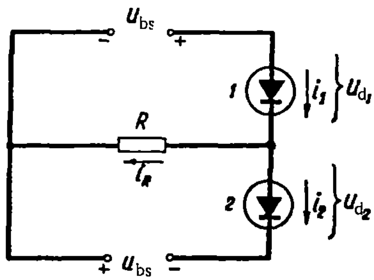

FIGURE 2.55. Twin circuit.

If now a positive voltage is applied to input A, which increases the diode voltage to u_C, the diode current drops abruptly from i_A to i_C. When the input voltage is lifted, the diode D is cut off and the tunnel diode continues to operate at point C. To sum up, when a signal is applied to input A, an output signal is produced at C, which persists after the input signal has been lifted. The output signal can be cleared by switching off the supply voltage (push button K). As u_{bs} increases after switching on again, the tunnel diode again operates at point A (Figure 2.52). The circuit shown in Figure 2.54 is thus an equivalent of the self-holding relay.

The output signal can also be cleared by applying a negative-voltage pulse to input B. To reset the tunnel diode, the voltage across R (Figure 2.52) must be equal to $u_{bs_1} - u_A$, since in this case the diode voltage is u_A and A

is therefore the operating point. When signal B is lifted, the tunnel diode remains at point A and no output signal is produced (the output is cleared to zero).

With two inputs (A and B), the circuit in Figure 2.54 is equivalent to the bistable circuit in Figure 2.25, designed of four logic elements.

The twin circuit is another logic element designed with tunnel diodes (Figure 2.55). In this circuit two diodes and two equal bias-voltage sources (u_{bs}) are connected in series. Let the twin circuit consist of two matched diodes, whose current-voltage characteristics are shown in Figure 2.43. Given the diode characteristics, the current-voltage characteristics of all the components in the twin circuit can be plotted. The resistance of R is generally much higher than the resistance of the diodes. Therefore, we can neglect i_R in comparison with the diode currents (i_1 and i_2) and we may set $i_R=0$ and $i_1=i_2$. Since now the same current flows through each diode, the current-voltage characteristic of the twin circuit can be constructed by summing the diode voltages:

$$2\,u_{bs}=u_{d_1}+u_{d_2},$$

where u_{d_1}, u_{d_2} = the voltages across the first and the second diodes;
$2\,u_{bs}$ = the bias voltage for a given current.

From Figure 2.43, it can be seen that when the current lies between the values $i_{max}>i>i_{min}$, the diode may have three different voltages, depending on whether it is operating in region 1, 2, or 3. On the same current range, these voltages give, in a turn circuit, six possible combinations 1—1, 1—2 1—3, 2—2, 2—3, and 3—3, i.e., there are six different voltage values to each value of current (Figure 2.56). Where $i<i_{min}$ or $i>i_{max}$, the voltage can assume only one value for each value of current.

The characteristic of a twin circuit is shown in Figure 2.56. As previously explained, this curve has six different regions for currents in the $i_{max}>i>i_{min}$ range. Five of these portions (1—1, 1—2, 1—3, 2—3, and 3—3) are stable, whereas the sixth (2—2) is unstable. This will now be considered in greater detail.

The following relationships describe a twin circuit (Figure 2.55):

$$u_{d_1}=u_{bs}-Ri_R; \tag{2.36}$$

$$u_{d_2}=u_{bs}+Ri_R; \tag{2.37}$$

$$i_R=i_1-i_2. \tag{2.38}$$

Let the two diodes operate in the second region of their respective characteristics. Since the same current flows through the two diodes, the diode voltages are equal. The voltage across R is zero, so that $i_R=0$. If, however, some random change in the tunnel-diode parameters or in the bias voltage u_{bs} produces a finite current, the following occurs.

When i_R appears, the voltage on the first diode decreases (2.36) and the voltage on the second diode increases (2.37). Since each diode is operating in the second region of its respective characteristic (Figure 2.43), the decrease in voltage on the first diode produces an increase in current i_1,

and the increase in voltage on the second diode produces a decrease in current i_2. On the other hand, the increase in i_1 and the decrease in i_2 will increase (2.38) the current i_R. The increase in i_R, in its turn, will further decrease the voltage on the first diode and increase the voltage on the second diode, etc. We thus have a cumulative process which will transfer the twin circuit from region 2—2 to one of the stable regions (1—2, 1—3, or 2—3).

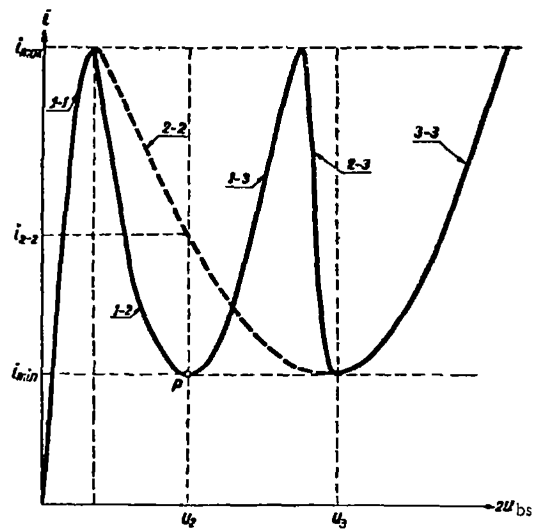

FIGURE 2.56. Current-voltage characteristics of two tunnel diodes connected in series.

Suppose that in Figure 2.56 the sum of the two bias voltages applied to the twin circuit is equal to u_2. Since the range 2—2 is unstable, the circuit will settle at point P. This point is obtained by adding the voltages across the diodes when the current is equal to i_{min}. We shall assume that one diode is operating at point B, and the other at point C of Figure 2.43. The sum of these voltages is equal to u_2. The value of 1 is assigned to u_B, and the value of 0 to u_C.

Since the diode characteristics are identical, either diode can acquire voltage u_B or u_C when the bias voltage is applied to the circuit. However, having acquired one of the voltages, the diode will persist in its state indefinitely. This property of the twin circuit can be utilized in various logic elements.

For example, Figure 2.57 shows a bistable circuit designed on the principle of the twin circuit. This circuit performs the same logical operations as that shown in Figure 2.25. When input A is energized, output C is set to 1. In this case the voltage on the second diode is u_B and that on the first diode is u_C. The output remains at 1 after the input A has been de-energized. If now input B is energized, the voltage on the second diode decreases sharply and that on the first diode increases. The output voltage drops to 0; the diode D preventing reversal in its polarity.

The twin circuit is also suitable for designing a static self-holding switch This switching circuit is shown in Figure 2.58. If a voltage is applied to the

input (A) of this switch, output C is set to 1 (voltage u_B, Figure 2.43) and output B is cleared to 0 (voltage u_C). When push button K is pressed, the voltage on the second diode drops sharply, and that on the first diode increases. The 1 is thus transferred from output C to output B. When input A is again energized, output C is reset to 1. In terms of logical operations, this circuit is analogous to those shown in Figure 2.40.

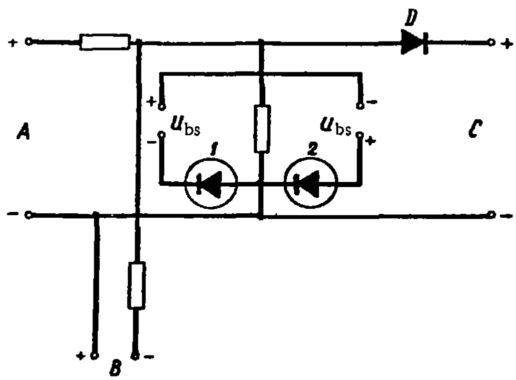

FIGURE 2.57. Schematic diagram of tunnel-diode bistable circuit.

Note that when the bias voltage u_{bs} is first applied to the circuit in Figure 2.58, we cannot state a priori which of the outputs will be set to 1. To eliminate this ambiguity we must press push button K. This will set output B to 1 and clear output C to 0.

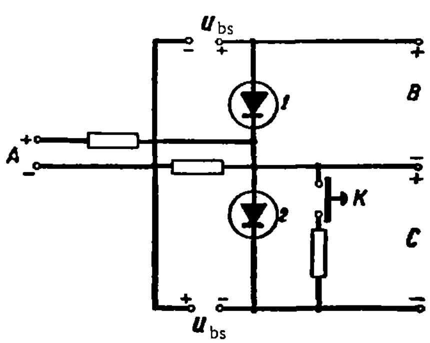

FIGURE 2.58. Static switching circuit.

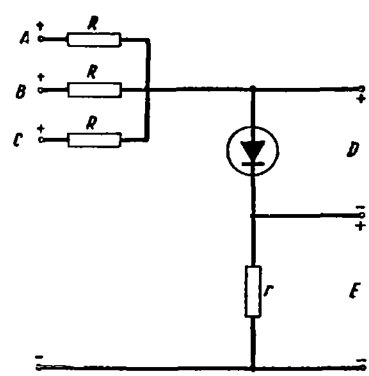

FIGURE 2.59. Tunnel-diode one-bit adder.

The tunnel diode can be used in a very simple circuit (Figure 2.59) as a one-digit binary adder for three numbers (two addends and carry from a lower position). This circuit functions as follows.

The resistors R and r are so chosen that when only one of the circuit inputs (A, B, or C) is energized the tunnel-diode voltage is u_{d_F} (Figure 2.60).

In this case the diode operates at point F, transmitting current i_b.

When two equal voltages are impressed at any two inputs, the diode voltage increases and its operating point shifts to G, and the current decreases to i_a. If all three inputs are energized, the diode voltage increases further and the diode current rises again to i_b (point H).

Let us assume that the output currents (D and E) are negligible. The current passing through r is then equal to the diode current (i_d). The voltage across r is therefore equal to

$$u_r = i_d r.$$

Hence, since r = constant, the output voltage E is proportional to the tunnel-diode current (i_d). If the voltage $i_b r$ is now coded as 1, and $i_a r$ as 0, the output E is set to 1 when (Figure 2.60) one input or three inputs are energized. Referring to Figure 2.12, it can be seen that when output E is set to 1, it represents the sum signal.

Let now the voltage u_{d_F} represent signal 0, and the voltages u_{d_G} and u_{d_H} signal 1. Output D is then set to 1 when two or three inputs are energized. Hence (Figure 2.12), when output D is set, it represents the carry signal.

We now calculate the resistors R and r. The following equations describe the adder circuit (Figure 2.59):

$$u = (R + r) i_1 + u_{d_F}; \tag{2.39}$$

$$u = \left(\frac{R}{2} + r\right) i_2 + u_{d_G}; \tag{2.40}$$

$$u = \left(\frac{R}{3} + r\right) i_3 + u_{d_H}, \tag{2.41}$$

where u = the voltage applied to one or several inputs of the circuit;

i_1, u_{d_F} = the current and the voltage of the tunnel diode when voltage u is applied to one of the circuit inputs;

i_2, u_{d_G} = the current and voltage when any two inputs are energized;

i_3, u_{d_H} = the current and voltage when all three inputs are energized.

Seeing that at points F and H (Figurc 2.60)

$$i_1 = i_2 = i_b,$$

we derive from (2.39) and (2.41)

$$R = \frac{3}{2} \frac{u_{d_H} - u_{d_F}}{i_b}. \tag{2.42}$$

Therefore, reading the voltages u_{d_F} and u_{d_H} off the tunnel-diode characteristic (Figure 2.60), we may determine the resistance R.

Inserting R from (2.42) into (2.40) and remembering that $i_2 = i_a$ we have

$$r = \frac{u - u_{d_G}}{i_a} - \frac{3}{4} \cdot \frac{u_{d_H} - u_{d_F}}{i_b}. \tag{2.43}$$

Given the parameters u, u_{d_G} and i_a, we derive from (2.43) the unknown resistance r.

The twin circuit can also be used to design AND, OR, and NOT gates. These are, however, rather complicated and are employed only when exceptionally high-speed switching is required.

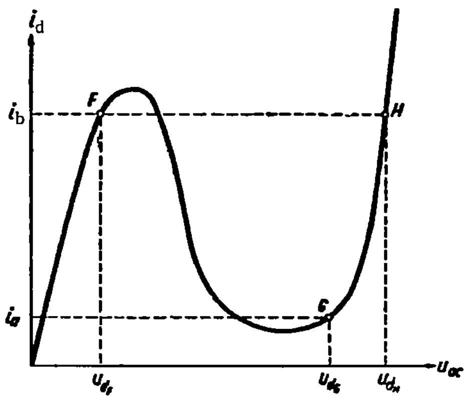

FIGURE 2.60. Operating points of a tunnel diode.

8. LOGICAL OPERATIONS WITH ANALOG QUANTITIES

Electronic logic elements are widely used not only in digital circuits which transmit discrete signals, but also to process information transmitted by continuously varying signals (analog quantities). In the latter case the logic element must choose not between two values 1 and 0, but rather between an infinite number of values of a smoothly varying signal. Therefore, the requirements of these logic elements are far more stringent as regards accuracy and stability.

Before discussing some examples of the application of electronic logic elements, we shall first consider an important property of the OR gate, when a continuous voltage is applied to its inputs.

In Figure 2.61, three continous voltages u_1, u_2, and u_3 are applied simultaneously to the input of an OR gate. At a given instant, u_1 is greater than u_2 and u_3. Voltage u_1 causes a current i_1 to flow through R, producing across it a voltage drop $u_R = Ri_1 = u_1$. Since u_1 is greater than u_2 and u_3, diodes 2 and

3 are cut off and no current flows in them. Current flows only through the diode to which the highest voltage is applied. If signal lights are connected to the circuits of the three diodes only the light in branch 3 will be energized. The resistance of the signal lights should be one or two orders of magnitude less than the resistance of R. Otherwise, the lights will introduce considerable distortion in the circuit s operation.

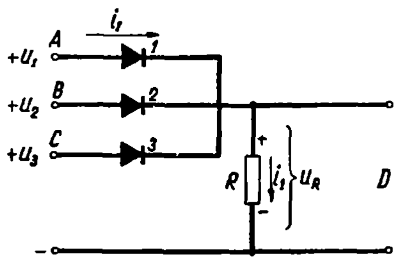

FIGURE 2.61. Determination of maximum voltage by means of an OR gate.

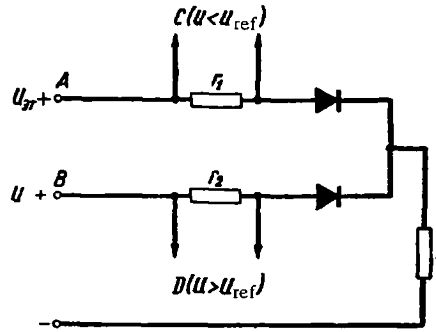

FIGURE 2.62. Comparator circuit for two voltages.

Therefore, the voltage at the output of the OR circuit (D) is equal to the highest of the voltages applied to the input. The OR circuit thus selects the highest of several input voltages, measures it, and indicates the branch to which this voltage is applied.

Let us now consider a two-input OR circuit (Figure 2.62). A reference voltage u_{ref} is applied to input A, and a variable voltage u is applied to input B. Current passes through r_1 only if the measured voltage is less than the reference voltage. Consequently, a signal appears at output C only when $u<u_{ref}$. Similarly, a signal is produced at output D only when $u>u_{ref}$. This circuit thus compares a given continuous voltage with a reference voltage.

Some examples in the application of logic in continuous current circuits will now be considered.

a) Determination of the largest current from three motors connected in parallel

Equal resistors are introduced into the armature circuit of each motor (Figure 2.63). The voltages across these resistors are proportional to the motor currents. These three resistances are coupled to an OR circuit. A voltmeter provided at the output of this circuit will therefore measure the highest input voltage. If the voltmeter is graduated in amperes, it will give the largest current. A signal light will indicate the circuit whose current is being measured at the moment.

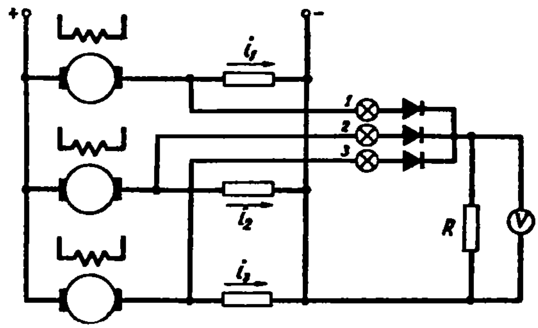

FIGURE 2.63. Circuit for measuring the largest current of three motors.

b) Automatic starting of a d-c motor

Consider the circuit in Figure 2.64 where a d-c motor is started by the successive shorting of two resistors connected in the armature circuit. When contact 1 is made, maximum armature current flows (point *a*, Figure 2.65). As the speed of the motor increases, this current decreases and when it drops to i_{min} (point *b*) contact 2 is made and the current jumps to point *c*; when the current again drops to i_{min} (point *d*), contact 3, is made.

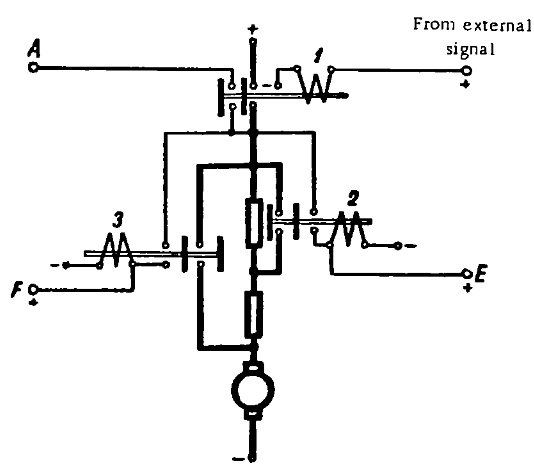

FIGURE 2.64. Motor-starting circuit.

Contact 1 is made by an externally supplied signal. Contacts 2 and 3 are made according to the following logical specifications:

a) if contact 1 is made and $i \leqslant i_{min}$, make contact 2;

b) if the motor voltage reaches u_{min} as the motor speeds up and $i \leqslant i_{min}$, make contact 3;

c) if contact 1 is broken, break contacts 2 and 3.

The last condition is essential for resetting the circuit for the next starting of the motor.

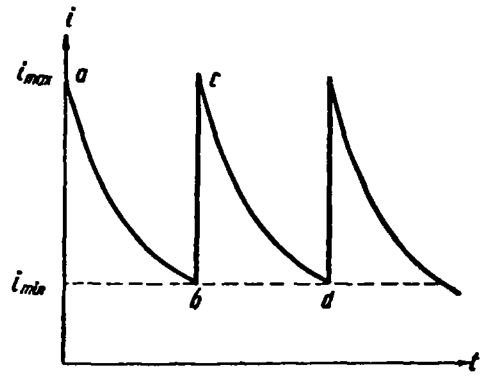

FIGURE 2.65. Diagram of starting currents.

The logic system fulfilling these specifications is shown in Figure 2.66. The system has four inputs. Input A receives the signal which makes contact 1 in Figure 2.64. A voltage u which is proportional to the armature current is applied to input B of a current comparator designed according to the circuit shown in Figure 2.62. The reference voltage (u_{ref}) is applied to input C, and the motor voltage (u_m) to input D. The OR_1 gate produces an output signal when the motor current drops to $i \leqslant i_{min}$. The OR_2 gate acting as a comparator produces an output signal when the voltage of the motor speeding up exceeds u_{min}.

The system operates as follows; when contact 1 is made, input A is set to 1. When the motor current decreases to i_{min}, the output of the OR_1 gate is also set to 1. The AND_1 gate now produces an output signal which makes contact 2 (Figure 2.64). When contact 2 is made, the motor current increases abruptly and then starts decreasing. The motor voltage continues to increase and when inequalities $i<i_{min}$ and $u_m>u_{min}$ are satisfied, the OR_2 gate is set to 1 and contact 3 is made.

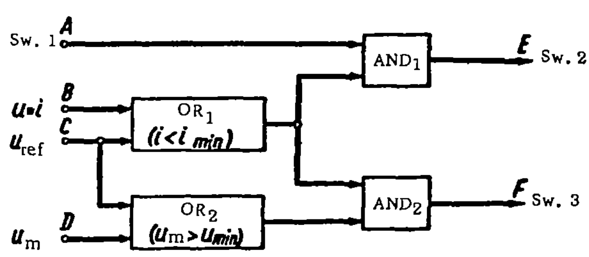

FIGURE 2.66. Logic system for motor starting.

When contact 1 is broken, the coils of contacts 2 and 3 and the system in Figure 2.66 are de-energized, and the circuit of Figure 2.64 is reset for further operation.

c) A two-mode automatic regulator for battery charging

One of the optimum methods of battery charging is to charge the battery with a constant current while the battery voltage gradually increases to its maximum value, and then to switch over to constant-voltage charging (Figure 2.67).

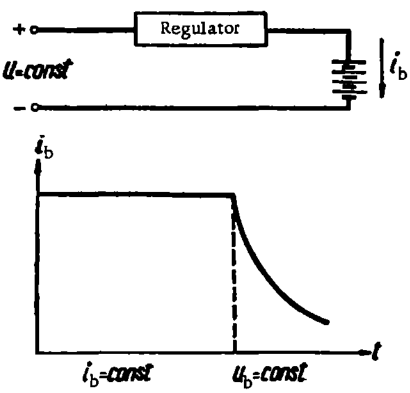

FIGURE 2.67. A battery-charging circuit and charging characteristics.

This switching can be performed by a simple circuit consisting of two gates (Figure 2.68). The circuit has two inputs, one energized by the battery voltage (u_b) and the other by a reference voltage (u_{ref}). When the battery voltage rises to its maximum value, the comparator produces a signal (*C*) which switches the regulator over to the constant-voltage mode. The signal (*D*) which switches the regulator over to the constant-current mode is simultaneously cleared.

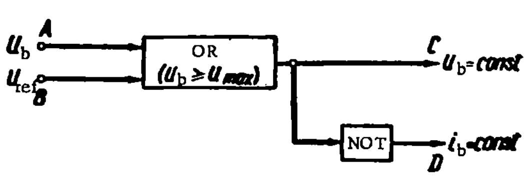

FIGURE 2.68. Logic system for switching charging modes.

Most analog computer circuits employ electronic amplifiers. These amplifiers mostly use transistors or triodes (Figure 2.69). The symbol for an amplifier is a rectangle with an arrow pointing in the direction of amplification (Figure 2.70). These amplifiers amplify the input voltage u by a factor of k and simultaneously (Figure 2.69) change its sign.

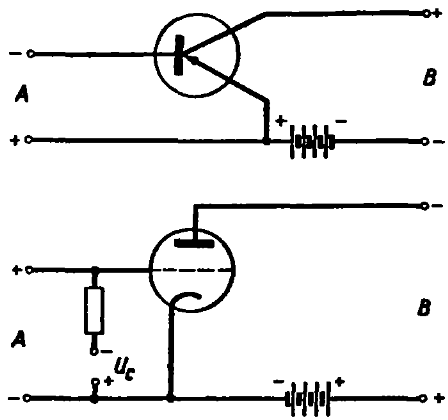

FIGURE 2.69. Transistor and triode amplifiers.

Computer circuits often require very high (10^6 and higher) gain, and to accomplish this, several (generally three) amplifiers are connected in series. In this case, an amplifier stage is also represented by a rectangle, as shown in Figure 2.70.

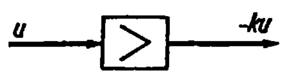

FIGURE 2.70. Symbolic representation of an amplifier.

Let us consider how various operations with analog voltages are performed.

a) Operational amplifiers

To achieve the highest accuracy in a computing circuit, the maximum value of the signal voltages should be as close as possible to the nominal voltage of the circuit. The maximum voltage values encountered in practice, however, may assume arbitrary values. The amplifiers in Figure 2.69 are inadequate for adapting these voltages, since they have a constant gain factor.

Consequently, computing circuits often make use of **operational amplifiers** in which the gain may be easily and smoothly regulated. The block diagram of an operational amplifier is shown in Figure 2.71. We shall now determine the gain of this amplifier.

$$k_g = \frac{u_{out}}{u_{in}}.$$

Since the current i_0 (grid current in a tube or base current in a transistor) is many times smaller than the currents i_1 and i, we shall neglect it, setting $i_0=0$. We therefore have the following relationships for the circuit in Figure 2.71:

$$\left.\begin{aligned} &i_1+i=0;\\ &u_{in}=i_1R_1+u_A;\\ &u_{out}=iR+u_A;\\ &u_{out}=-ku_A,\end{aligned}\right\} \tag{2.44}$$

where k is the amplifier gain:

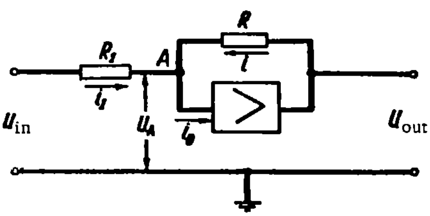

FIGURE 2.71. Block diagram of an operational amplifier.

Solving the simultaneous equations (2.44), we derive

$$k_g = \frac{u_{out}}{u_{in}} = -\frac{R}{R_1 + \frac{R_1+R}{k}}.$$

As stated previously, the amplifier gain (k) is generally very high. Therefore, setting ($k=\infty$), we obtain

$$k_g = \frac{u_{out}}{u_{in}} = -\frac{R}{R_1}. \tag{2.45}$$

Thus, the gain of the operational amplifier can be regulated by varying R or R_1.

b) Summing of voltages

Consider the circuit in Figure 2.72. As before, we take $i_0=0$. Therefore,

$$i+i_1+i_2+\ldots+i_n=0;$$
$$u_{\text{in}_1}=i_1R_1+u_A;$$
$$u_{\text{in}_2}=i_2R_2+u_A;$$
$$\cdots\cdots\cdots$$
$$u_{\text{in}_n}=i_nR_n+u_A;$$
$$u_{\text{out}}=iR+u_A;$$
$$u_{\text{out}}=-ku_A.$$

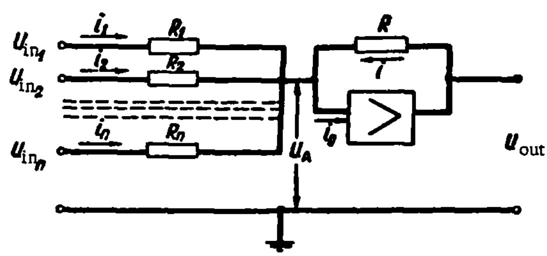

FIGURE 2.72. The summing amplifier.

Solving these simultaneous equations, we derive

$$\left[1+k+\left(\frac{1}{R_1}+\frac{1}{R_2}+\ldots+\frac{1}{R_n}\right)R\right]u_{\text{out}}=$$
$$=-\left(\frac{u_{\text{in}_1}}{R_1}+\frac{u_{\text{in}_2}}{R_2}+\ldots+\frac{u_{\text{in}_n}}{R_n}\right)kR$$

or

$$\left(1+k+R\sum_{m=1}^{n}\frac{1}{R_m}\right)u_{\text{out}}=-kR\sum_{m=1}^{n}\frac{u_{\text{in}_m}}{R_m},$$

whence

$$u_{\text{out}}=-\frac{R\sum_{m=1}^{n}\frac{u_{\text{in}_m}}{R_m}}{1+\frac{1}{k}+\frac{R}{k}\sum_{m=1}^{n}\frac{1}{R_m}}.$$

Setting $k=\infty$, we obtain

$$u_{\text{out}}=-R\sum_{m=1}^{n}\frac{u_{\text{in}_m}}{R_m}. \qquad (2.46)$$

If now $R=R_1=R_2=\ldots=R_n$, then

$$u_{\text{out}}=-\sum_{m=1}^{n}u_{\text{in}_m}. \qquad (2.47)$$

The circuit in Figure 2.72 thus sums any given number of voltages, and is called a summing amplifier.

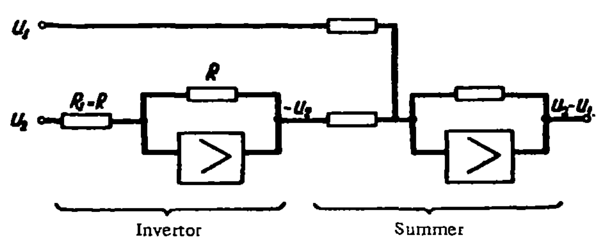

FIGURE 2.73. Block-diagram for voltage subtraction.

c) Subtraction of voltages

Subtraction is readily reducible to addition:

$$u_2 - u_1 = -[u_1 + (-u_2)].$$

The diminued u_2, is applied to an operational amplifier with a gain of 1 ($R_1 = R$) (Figure 2.73). This amplifier inverts the sign of u_2, which is then fed into a summer together with the subtrahend voltage.

d) Voltage integration

Voltage integration is performed by the circuit shown in Figure 2.74.

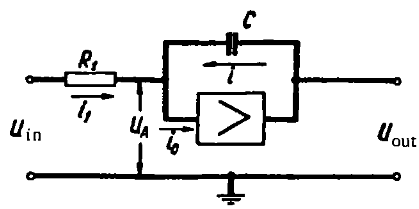

FIGURE 2.74. Voltage integrator.

Neglecting i_0, we have for this circuit

$$i = -i_1;$$

$$u_{in} = i_1 R_1 + u_A;$$

$$u_{out} = \frac{1}{C}\int i\,dt + u_A;$$

$$u_{out} = -k u_A.$$

Solving these equations, we obtain

$$u_{out}\left(1 + \frac{1}{k}\right) = -\frac{1}{C}\int \frac{u_{in} + \frac{1}{k} u_{out}}{R_1} dt.$$

Setting $k = \infty$, we have

$$u_{out} = -\frac{1}{CR_1}\int u_{in}\,dt. \tag{2.48}$$

Integration is often performed with nonzero initial conditions. Special circuits must be provided in integrators to set these conditions. This setting is carried out by charging a capacitor to the voltage specified. The capacitor, as a rule, must not be disconnected from the integrator feedback circuit, since the open feedback loop would result in instability and in error of solution. The capacitor is charged from an external source (u_C) as shown in Figure 2.75. The upper position of the switch in the circuit corresponds to the initial condition mode, and the lower position to the operate mode of the circuit.

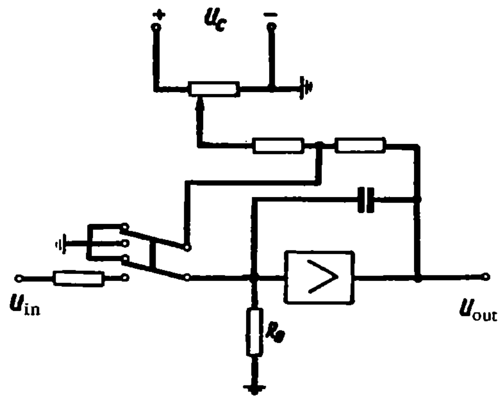

FIGURE 2.75. Initial condition and operate circuit for an integrator amplifier.

e) Voltage differentiation

Voltage is differentiated by the circuit in Figure 2.76. In this circuit, $i_0=0$,

$$i=-i_1;$$

$$u_{in}=\frac{1}{C}\int i_1 dt+u_A;$$

$$u_{out}=iR+u_A;$$

$$u_{out}=-ku_A.$$

Solving these equations simultaneously, we obtain

$$u_{in}+\frac{1}{k}u_{out}=-\frac{1}{CR}\int\left(1+\frac{1}{k}\right)u_{out}dt.$$

Setting $k=\infty$, we obtain

$$u_{\text{in}}=-\frac{1}{CR}\int u_{\text{out}}dt$$

or

$$u_{\text{out}}=-CR\frac{du_{\text{in}}}{dt}+A, \tag{2.49}$$

where A is the integration constant specified by additional conditions.

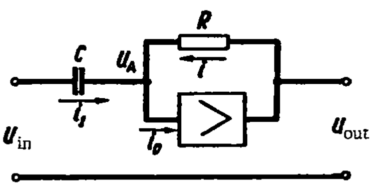

FIGURE 2.76. Voltage differentiator.

Differentiation, unlike integration, has the following serious shortcomings:

a) the differentiator (Figure 2.76) amplifies the high-frequency noise, thus increasing the error of differentiation;

b) the increased sensitivity of the circuit to high-frequency harmonics may induce high-frequency oscillations in the circuit.

Consequently, differentiation is performed only in extreme cases. Whenever possible, the order of the differential equation must be lowered without resorting to differentiation.

For example, find the current in the equation

$$u=iR+L\frac{di}{dt},$$

where R and L are constants.

Integrating the two sides of the equation, gives

$$\int udt=R\int idt+Li+A,$$

where A is the constant of integration.

Here we have substituted integration for differentiation (the technique for the solution of these equations will be illustrated at the end of the chapter).

f) Coefficient-variation uits

Coefficient-variation units are used when some coefficients in the equations must vary with time during computation.

A coefficient-variation unit using a programmed switch selects at definite time intervals (Δt) consecutive positions which connect different voltages (Figure 2.77). These voltages are chosen so as to ensure a satisfactory approximation of the given time-variation curve of the coefficient (the dashed curve in Figure 2.77).

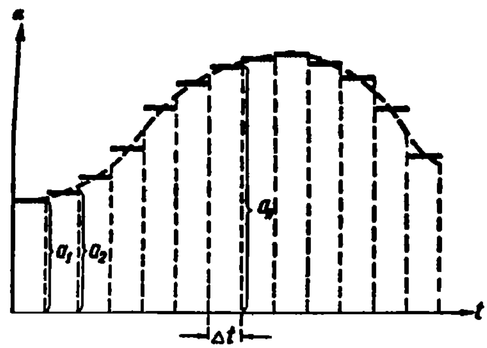

FIGURE 2.77. Voltages set up by the coefficient-variation unit.

g) Nonlinear function generators

Nonlinear function generators are used when a nonlinear function $u_{out}=f(u_{in})$, defined graphically, is involved in the computations. In order to generate the nonlinear function, it must be approximated by straight segments (Figure 2.78) and the breakpoints a_2, a_3... and b_1, b_2, b_3... must be determined.

The simplest function generator is shown in Figure 2.79. It consists of several circuits, each corresponding to a definite segment of the approximated function. For example, the circuit with the resistor R_1 corresponds to segment (I) of the function (Figure 2.78); the circuit with R_2 and e_2 to segment (II), etc. The switching from one circuit to another is achieved by a switch operated by the input voltage: if $u_{in}<u_{in_{1-2}}$ (Figure 2.78), the switch is in the first position; when $u_{in_{2-3}}>u_{in}>u_{in_{1-2}}$, the switch connects the second circuit, etc.

We shall now determine the circuit parameters of this function generator.

For the n-th position of the switch (Figure 2.79), we have

$$u_{in}=e_n+R_n i_n+u_{out};$$

$$u_{out}=ri_n.$$

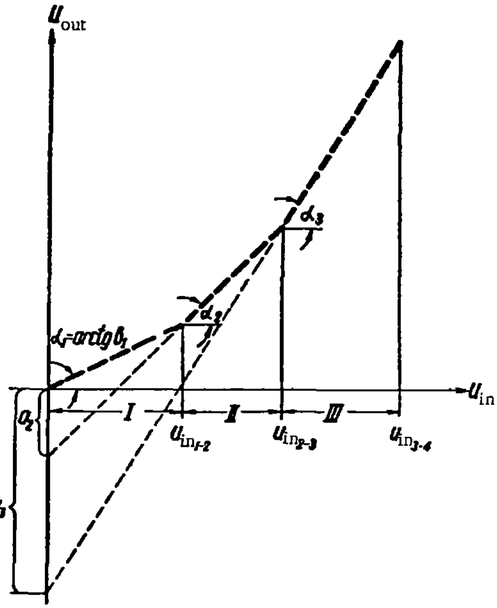

FIGURE 2.78. Piecewise linear approximation of a function.

Solving these simultaneous equations gives

$$u_{out}=\frac{u_{in}-e_n}{1+\frac{R_n}{r}}. \tag{2.50}$$

On the other hand, for the n-th segment of the piecewise approximation (Figure 2.78),

$$u_{out}=-a_n+b_n u_{in}. \tag{2.51}$$

Solving (2.50) and (2.51) simultaneous[1], we obtain

$$-a_n+b_n u_{in}=-\frac{r}{R_n+r}e_n+\frac{r}{R_n+r}u_{in}.$$

The parameters of the n-th circuit of the nonlinear unit are thus determined by the relationships

$$e_n=\frac{R_n+r}{r}a_n; \tag{2.52}$$

$$R_n=\frac{1-b_n}{b_n}r. \tag{2.53}$$

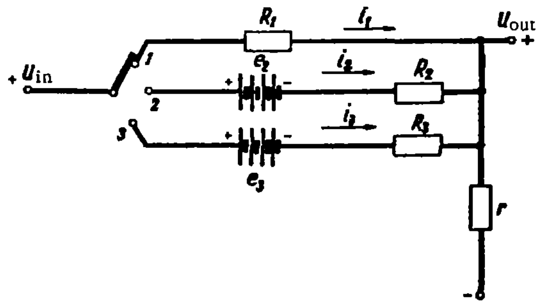

FIGURE 2.79. A simple function generator.

The resistance of r is arbitrary, but it must be much smaller than the resistance of R_1, R_2...

This function generator has a switch actuated by the input voltage. This switch can be replaced by diodes. For example, Figure 2.80 shows a unit consisting of three resistance circuits, R_1, R_2, and R_3. Such a unit can be designed for any number of circuits. The number of these circuits is determined by the number of segments in the piecewise linear approximation (Figure 2.78).

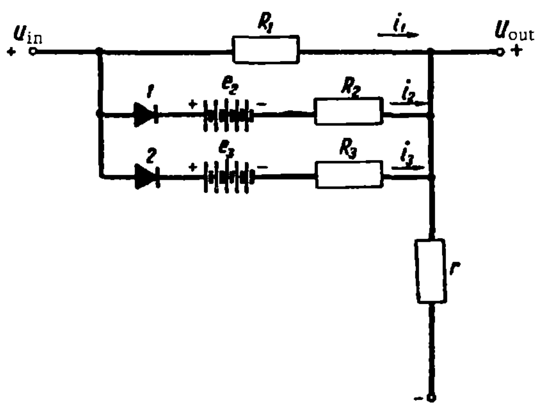

FIGURE 2.80. Static function generator.

The circuit shown in Figure 2.80 functions as follows. If the input voltage is relatively small ($u_{in} < u_{in_{1-2}}$ Figure 2.78), the emf sources (e_2 and e_3) cut off the corresponding diodes and current flows only through R_1. When the voltage u_{in} is greater than $u_{in_{1-2}}$, diode 1 conducts and current now flows through two circuits (R_1 and R_2). When the voltage rises to $u_{in_{2-3}}$, diode 2 conducts and current passes through all the three circuits.

The circuit parameters of this function generator are determined as follows.

When $u_{in} \leqslant u_{in_{1-2}}$,

$$u_{in} = R_1 i_1 + u_{out};$$

$$u_{out} = r i_1.$$

Hence

$$u_{out} = \frac{r}{R_1 + r} u_{in}.$$

The resistor r is taken so small as to be negligible in comparison with the other resistors R_1, R_2, and R_3. Then

$$u_{out} = \frac{r}{R_1} u_{in}.$$

It can be seen from Figure 2.78 that for the first portion

$$u_{out} = b_1 u_{in}. \tag{2.54}$$

We therefore have from the last two equations

$$R_1 = \frac{r}{b_1}. \tag{2.55}$$

If $u_{in_{2-3}} \geqslant u_{in} \geqslant u_{in_{1-2}}$, the voltages in the function generator are given by the following equations (r is neglected in comparison with the other resistances):

$$u_{in} = R_1 i_1;$$

$$u_{in} = e_2 + R_2 i_2;$$

$$u_{out} = (i_1 + i_2) r.$$

Solving these equations simultaneously, gives

$$u_{out} = -\frac{r}{R_2} e_2 + \frac{(R_1 + R_2) r}{R_1 R_2} u_{in}. \tag{2.56}$$

The second segment of the piecewise curve in Figure 2.78 is defined by the equation

$$u_{out} = -a_2 + b_2 u_{in}.$$

Equating the right-hand sides of the last two equations term by term gives the parameters

$$e_2 = \frac{R_2}{r} a_2;$$

$$R_2 = \frac{R_1 r}{b_2 R_1 - r}.$$

Solving these equations simultaneously with (2. 55), gives

$$e_2 = \frac{a_2}{b_2 - b_1}, \tag{2.57}$$

$$R_2 = \frac{r}{b_2 - b_1}. \tag{2.58}$$

From these equations it can be seen that it is essential, when approximating, that $b_2 > b_1$. This condition in Figure 2. 78 is always satisfied.

We have previously said that diode 1 (Figure 2. 80) conducts when

$$u_{\text{in}} = u_{\text{in}_{1-2}}.$$

The following equality must therefore be satisfied (the resistance of r is neglected):

$$e_2 = u_{\text{in}_{1-2}}. \tag{2.59}$$

We shall verify this equality using the parameters given by (2. 55), (2. 57), and (2. 59).

It follows from (2. 59) that

$$u_{\text{out}_{1-2}} = b_1 u_{\text{in}_{1-2}},$$

where $u_{\text{out}_{1-2}}$ is the output voltage corresponding to input voltage $u_{\text{in}_{1-2}}$.

On the other hand, from (2. 56),

$$u_{\text{out}_{1-2}} = -\frac{r}{R_2} e_2 + \frac{(R_1 + R_2) r}{R_1 R_2} u_{\text{in}_{1-2}}.$$

Eliminating $u_{\text{out}_{1-2}}$ from these two equations gives the input voltage

$$u_{\text{in}_{1-2}} = \frac{R_1 r}{(R_1 + R_2) r - b_1 R_1 R_2} e_2.$$

Inserting R_1 from (2.55) and R_2 from (2.58), gives

$$u_{\text{in}_{1-2}} = e_2.$$

Condition (2.59) is thus satisfied.

The parameters of the third circuit (Figure 2.80) are determined similarly. Since the resistance r is taken very small, the output voltage from the function generator is not high, and can be fed directly to an amplifier.

h) Multiplication of two voltages

Multiplication is reduced to addition and squaring. This reduction is based on the equality

$$(u_1+u_2)^2-(u_1-u_2)^2=4u_1u_2. \tag{2.60}$$

To prove the validity of this equality, it suffices to expand the parentheses in the left-hand side.

A circuit for multiplying two voltages (u_1 and u_2) is shown in Figure 2.81 and functions as follows.

Amplifier 2 adds the voltages (u_1+u_2) and inverts the sign of the sum. Amplifier 1 inverts the sign of voltage u_2; the output of amplifier 3 is therefore the difference (u_1-u_2) with opposite sign. The two function generators give a piecewise linear approximation of the quadratic function $u_{\text{out}}=u^2_{\text{in}}$. The outputs of these generators are, respectively, the square of the sum and the square of the difference of the two voltages. Amplifier 4 changes the sign of the square of the difference, and amplifier 5 adds $[(u_1+u_2)^2]+$ $+[-(u_1-u_2)^2]$ and scales the sum (reducing it to 1/4), so that its output (see equation (2.60)) is the product of the two voltages.

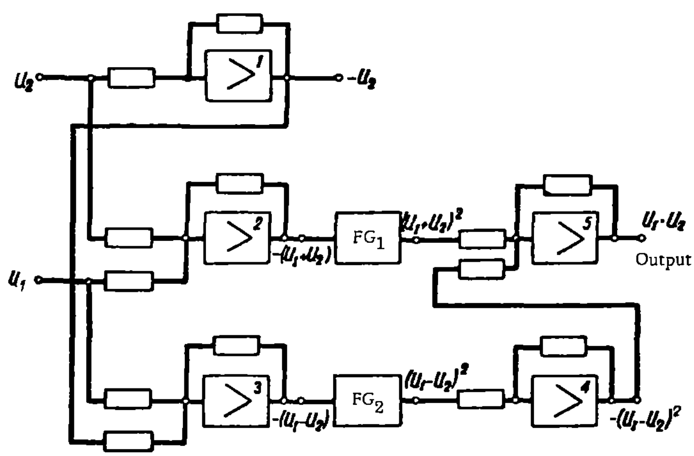

FIGURE 2.81. Circuit for multiplication of two voltages [quarter square multiplier].

i) Division of two voltages

The division of two voltages is reduced to multiplication:

$$\frac{u_1}{u_2}=u_1\frac{1}{u_2}.$$

The factor $\frac{1}{u_2}$ is generated by a function generator programmed to approximate the function

$$u_{\text{out}}=\frac{1}{u_{\text{in}}}.$$

* *

In conclusion let us consider as an example the computation of the current curve $i=f(t)$, when the switch in the circuit of Figure 2.82 is closed. To simplify the discussion, we shall assume zero initial conditions.

The circuit is described by the equation

$$u=Ri+wS_{\text{c}}\frac{dB}{dt}\cdot 10^{-8}, \tag{2.61}$$

where w = the number of windings in the coil;
S_{c} = the cross-sectional area of the coil;
B = the flux density in the coil.

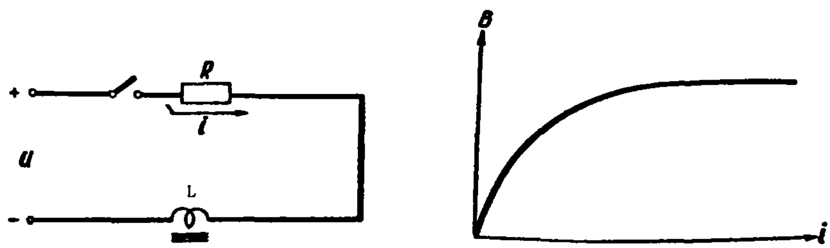

FIGURE 2.82. A nonlinear circuit consisting of a resistor and a coil.

The flux density in the coil is a nonlinear function of the coil current, and is graphically represented in Figure 2.82.

An initial examination of (2.61) seems to indicate that differentiation is involved. We have explained, however, that differentiation should be avoided as far as possible owing to the considerable errors inherent in this operation. We shall try to determine the current without differentiation.

We therefore rewrite (2.61) in the form:

$$\frac{dB}{dt}=\frac{u-Ri}{wS_c\cdot 10^{-8}}. \qquad (2.62)$$

Set

$$k_1=\frac{1}{wS_c\cdot 10^{-8}};$$

$$k_2=\frac{R}{wS_c\cdot 10^{-8}}.$$

Equation (2.62) takes on the simple form:

$$\frac{dB}{dt}=k_1u-k_2i. \qquad (2.63)$$

The block diagram for the solution of this equation is shown in Figure 2.83. The corresponding circuit is designed as follows.

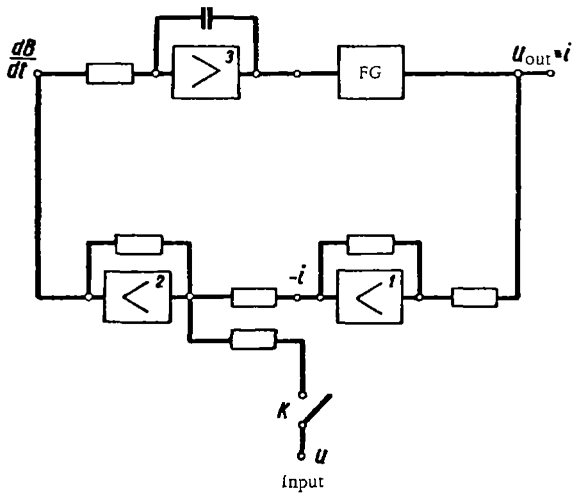

FIGURE 2.83. Block diagram for current computation.

We see from (2.63) that $\left(\frac{dB}{dt}\right)$ is equal to the difference of the voltage and the current in the circuit, each multiplied by a corresponding coefficient or, in other words, each taken to a different scale. We tentatively assume that the voltage u and the current i are known (the voltage is indeed known, but the current is to be determined).

We then feed the "known" current (more precisely, a voltage proportional to this current) to amplifier 1, which reverses the sign of the current. The inputs to amplifier 2 are the voltage u and the current with a minus sign, $-i$. Amplifier 2 therefore subtracts the current from the voltage. A suitable choice of resistors at the input of this amplifier scales the voltage and the current as required before subtraction. Hence, the output of amplifier 2 is a voltage $(u=k_1u-k_2i)$ which, according to (2.63), is equal to the derivative $\left(\frac{dB}{dt}\right)$. Feeding the derivative into the integrator (3) we obtain the flux density B. The flux density B is then fed into the function generator which is programmed to approximate the graphical function $i=\varphi(-B)$. The output of the integrator therefore gives the current i. This current is then fed to the input of amplifier 1.

When voltage u is applied to the input of the circuit, its output gives the required curve $i=f(t)$.

It follows from the preceding example that individual operations are reduced to the design of an electronic model (analog) of the electrical circuit in question. The computations carried out using these elements are therefore called analog computations.

Chapter III

INTRODUCTION TO THE THEORY OF PROBABILITY AND ELEMENTS OF INFORMATION THEORY

1. FUNDAMENTALS OF THE THEORY OF PROBABILITY

The behavior of any system subject to arbitrary external effects is determined by the laws governing the system. These laws, however, especially in cybernetic systems, are so complicated that their formulation in terms of differential equations is either impossible or impracticable. In this case it would be convenient to consider the behavior of the system as a random process, which can be analyzed by the theory of probability. This theory deals with functions of a random variable (or a variate) — a quantity which may assume one of many different values without giving any a priori indication as to exactly which of these values it will actually assume. If these values form a continuous function, the random variable is said to be continuous. A discrete random variable, on the contrary, can take on only a finite number of values.

Any of the values of a random variable can be considered as a random event, which may or may not occur during the period of observation.

Example. Consider the voltage output from one, two, or three 2-volt batteries connected together at random. The random variable (voltage) may take on the values of 2, 4, or 6 v. Any of these voltages, say 6 v, can be considered as a random event.

The probability of a random event is considered as a quantative estimate of its occurrence.

Probability theory does not enable us to predict the occurrence of an individual random event. If, however, after having made repeated measurements we can establish the behavior of the random variable in question, it is possible to determine the probability of occurrence of a random event.

Example. A locomotive leaves Riga on its way to the seashore. It is required to determine the distance it will cover (nonstop) consuming k kwhr of power.

This distance depends on the engineer's control, the voltage on the line, on the number of passengers, etc. It is therefore impossible to determine the distance for the locomotive on its first journey. The distance is a random variable, However, if the distance covered by the train is recorded on every trip, the probability that the train covers a preset distance consuming k kwhr can be determined .

The probability of a random event can be determined either by studying

the behavior of the event in question, or by processing data obtained in a series of experiments. Let us consider these possibilities in application to elementary examples.

A die is thrown on a horizontal plane. After a throw, the die will land with one of six numbers facing upward. Which face is upmost depends on the force of throw, the torque imparted to the die, and other factors which cannot be accounted for in advance. The landing of the die with any given number facing upward is therefore a random event.

The probability of this event can be determined as follows. If the mass of the die is homogeneous and its shape is regular (unbiased dice), the appearance of each of the six numbers is equiprobable. Therefore, setting the probability of a throw as 1, the probability of appearance of any predetermined number, say 3, is

$$P_3 = \frac{1}{6}.$$

Continuous observation of a random event can closely determine its probability. For example, in N tests, the event in question occurred n times. Taking the ratio

$$\mu = \frac{n}{N}, \tag{3.1}$$

we obtain a parameter called the *relative frequency* of the event.

The importance of this parameter is signified by the *law of large numbers*. This law states that in a sufficiently large number of measurements, the relative frequency of an event may be arbitrarily close to its probability (the proof is omitted). The probability of an event can thus be determined as the ratio of occurrences (n) to the number of tests (N), provided sufficient tests have been made.

For example, out of 55 measurements, the current amplitude dropped during a predetermined interval (from 1.2, a) 37 times. Determine the probability of a random measurement giving a current value in this interval.

In this example $N=55$, and $n=37$. The probability of the random event is therefore

$$(P)_{1,2>i>1} = \frac{37}{55} = 0.67.$$

The probability that an event *must* occur is taken as 1:

$$P = 1.$$

For example, when a die is thrown, one of the six numbers must face upward. The probability of an event that definitely will not occur is 0. Therefore, the probability of a random event always falls in the range

$$1 > P > 0. \tag{3.2}$$

We shall now consider a method for the determination of probabilities for compound events.

a) Probability of disjoint events

Events A and B are said to be disjoint (or mutually exclusive) if they cannot occur simultaneously. For example, when a die is tossed, two numbers cannot appear simultaneously on its upward face.

The probability of mutually exclusive events is equal to the sum of the probabilities of each event. For two events this is expressed by

$$P(A \vee B)=P(A)+P(B) \tag{3.3}$$

("the probability that either event A or event B occurs is equal to the sum of the probabilities of events A and B"). The proof of this equality follows from the definition of mutually exclusive events.

Since the logical OR is equivalent to logical addition, (3.3) can be written:

$$P(A+B)=P(A)+P(B). \tag{3.3a}$$

For example, when a die is thrown the outcome of 1 and 3 are mutually exclusive events, since these numbers cannot occur simultaneously. The probability of either 1 or 3 occurring when a die is tossed is thus equal to

$$P_{1+3}=P_1+P_3=\frac{1}{6}+\frac{1}{6}=\frac{1}{3}.$$

The probability of an event of whose outcome we are certain is set as 1. The sum of probabilities of all the possible mutually exclusive events constituting a random variable is thus also equal to one:

$$P_1+P_2+\ldots+P_m=1, \tag{3.4}$$

where P_1 = the probability of the first event;
P_2 = the probability of the second event;
....
P_m = the probability of the m-th event.

Since mutually exclusive events cannot occur simultaneously, the probability of the simultaneous occurrence (A and B) of two mutually exclusive events is therefore zero:

$$\left.\begin{array}{c} P(A \wedge B)=0; \\ \text{or} \\ P(AB)=0. \end{array}\right\} \tag{3.5}$$

b) Probability of joint events

J o i n t e v e n t s are events which may (though need not) occur simultaneously

For example, pick at random a chess piece. Consider two events: A - a white piece is chosen; B - a bishop is chosen. Since we may conceivably pick out a white bishop, the two events may occur simultaneously, i. e., they are joint. If, however, a white knight is picked out, only event A has occurred. Finally, if a black rook is selected, neither event A nor B have occurred.

Random events (both joint and disjoint) may be either d e p e n d e n t or i n d e p e n d e n t. Two events A and B are said to be i n d e p e n d e n t if the occurrence of one of them does not affect the probability of the other. Otherwise the events are d e p e n d e n t.

E x a m p l e s of independent and dependent events.

1. Two dice are tossed simultaneously. We are concerned with the simultaneous occurrence of the number 4 on both dice. The occurrence of this particular number on one die is independent of the number which occurs on the other die. The events in question are therefore i n d e p e n d e n t.

2. Three balls are placed in a box. Two of these are black and one is white. Consider the following two events: two balls are picked from the box one after the other. Event A indicates that the white ball has been picked first. Event B indicates that the white ball has been picked second. If the white ball has been selected first, the second ball must be black, i. e., in this case the probability of event B is zero. If the first ball is black, the second ball may be either black or white, i. e., the probability of event B is not zero. Hence, events A and B are d e p e n d e n t.

The probability of a dependent event (B) is called a c o n d i t i o n a l p r o b a b i l i t y, since it is conditional on the outcome of the other event (A). Conditional probability is denoted by

$$P(B/A).$$

We shall now consider the technique for the determination of the probability of two joint events.

In the example with the chess pieces, the following three classes of outcomes are possible in N tests:

1) event A occurs n_A times;
2) event B occurs n_B times;
3) both events A and B occur simultaneously (l times).

According to the law of large numbers, the probability of the simultaneous occurrence of events A and B is:

$$P(A \wedge B) = \frac{l}{N}.$$

Similarly the probability of occurrence of event A is equal to

$$P(A) = \frac{n_A}{N}.$$

The conditional probability of event B is the probability of the occurrence of this event subject to the condition that event A has occurred. Event A has occurred n_A times. Since out of these n_A times event B occurs only l times, we have

$$P(B/A) = \frac{l}{n_A}.$$

The probability of the simultaneous occurrence of the two dependent joint events A and B is equal to

$$\left.\begin{aligned} &P(A \wedge B) = \frac{l}{N} = \frac{n_A}{N} \cdot \frac{l}{n_A} = P(A)P(B/A); \\ &\text{or} \\ &P(AB) = P(A)P(B/A). \end{aligned}\right\} \qquad (3.6)$$

Therefore, from the definition of dependent events, the probability of the simultaneous occurrence of two independent joint events (3.6) is equal to

$$P(AB) = P(A)P(B). \qquad (3.6a)$$

Thus, we see that the probability of occurrence of only one of the two joint events (A or B, but not A and B) is equal to

$$P(A \vee B) = \frac{n_A + n_B - l}{N} = \frac{n_A}{N} + \frac{n_B}{N} - \frac{l}{N}.$$

In this equation the cases (l) when events A and B occur simultaneously are subtracted, since we are concerned with the probability of A or B only.

The probability of occurrence of one of two joint events is therefore given by the equation

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B).$$

Inserting the expression for $P(A \wedge B)$ from (3.6), we obtain for dependent events

$$\left.\begin{aligned} &P(A \vee B) = P(A) + P(B) - P(A)P(B/A) \\ &\text{or} \\ &P(A + B) = P(A) + P(B) - P(A)P(B/A). \end{aligned}\right\} \qquad (3.7)$$

For independent events (see (3.6a)), this equation takes on the form

$$P(A + B) = P(A) + P(B) - P(A)P(B). \qquad (3.7a)$$

Example 1. Two identical dice are tossed. Determine the probability of 1 appearing simultaneously on both dice.

When two dice are tossed, 1 may occur on one or both dice. The events in question are therefore joint. Since the probability of the occurrence of 1 on one die is independent of the number occurring on the other, the events in question are independent.

The probability of 1 appearing on the first die is

$$P(A)=\frac{1}{6},$$

and on the second die

$$P(B)=\frac{1}{6}.$$

The probability of the simultaneous occurrence of 1 on both dice is therefore from (3.6a) equal to

$$P(AB)=P(A)P(B)=\frac{1}{6}\cdot\frac{1}{6}=\frac{1}{36}.$$

Example 2. Determine the probability of the number 3 appearing on one die only, when two dice are tossed simultaneously.

Here, as before, the events are joint and independent.

The occurrence of a 3 on the first and on the second die is equiprobable:

$$P(A)=\frac{1}{6};\ P(B)=\frac{1}{6}.$$

The probability of a 3 occurring on one die only (and not on both dice simultaneously) is from (3.7a) equal to

$$P(A+B)=P(A)+P(B)-P(A)P(B)=\frac{1}{6}+\frac{1}{6}-\frac{1}{36}=\frac{11}{36}.$$

2. CHARACTERISTICS OF RANDOM VARIABLES

A random process is characterized by a *random function* described (Figure 3.1) by a family of different *sampling curves*.

Let us consider, for example, the voltage variation of a particular battery while being charged. The battery is charged by direct current at constant ambient temperature. No two batteries are perfectly alike. The characteristics of each battery depend on the composition and the active area

of its electrodes, the composition and volume of the electrolyte, heat transfer of the battery casing, and many other factors which cannot be accounted for with due accuracy. Battery charging is therefore a random process.

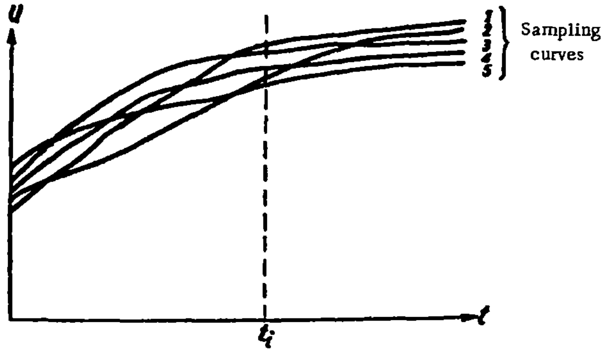

FIGURE 3.1. A random function of time.

If we record the battery voltage for different batteries during charging, a family of sampling curves of the random function is obtained (Figure 3.1). If we are concerned with the charging characteristics of a particular battery, the corresponding random function can be obtained by recording the results of several chargings of the given battery.

The random variable for a given instant t_i is determined by the intersection of a vertical line with the family of curves (Figure 3.1). This random variable, as we have already observed, consists of several random events (in Figure 3.1 there are five such events).

To determine a regular, nonrandom quantity, it suffices to obtain its characteristic value. For example, the statement "a nonrandum voltage at a given time instant (t_i) is equal to 5 v" fully characterizes a nonrandom (determinate) quantity (voltage in this case) at a given time instant.

The situation is different with random variables. A random variable is specified by its probability distribution, i.e., by the spectrum of all the values (e.g., voltages) which may occur at the point in question at a given time instant (t_i).

To specify a random function we must, moreover, know how the different random variables affect one another. This interaction of random variables known as correlation is considered in the next chapter.

Random processes studied by the theory of probability may be either stationary or nonstationary. We shall only consider stationary random processes whose probability functions are constant with time. Moreover, the theory of stationary random processes is comparatively simple and has been studied in sufficient detail, whereas an analysis of nonstationary processes is beyond the scope of this book.

Stationary random processes have the so called ergodic property. This highly important property amounts to the following: each individual sample of a random function taken over a sufficiently long period of time fully specifies, from the point of view of the theory of probability, the

entire family of sample curves of the random function. In this case, the analysis of the family of sample curves can be reduced to the analysis of a single curve. In stationary processes, the random variable is obtained not by examining a family of sample curves of the random function at a given time instant (Figure 3. 1), but rather by considering the behavior of a single sample along the time axis.

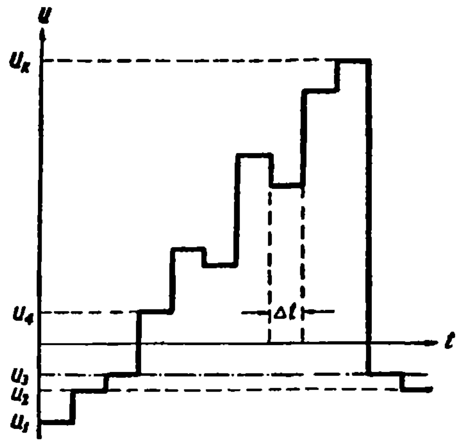

FIGURE 3. 2. Step random function.

Since in a stationary process a single sample specifies the entire random function, we shall speak in what follows of random function, omitting the qualifying word "sample".

It should be kept in mind that a random function is not "absolutely" random. In most cases we have some knowledge of its principal characteristic features. A random function is therefore not entirely arbitrary and is generally subject to certain constraints.

It is sometimes possible to determine a part of the frequency spectrum comprising the function; we can often specify the limiting values of the function or the range of its derivative. For each random function we therefore have a definite probability distribution of the random variables.

The random variables of a random function can either be discrete (Figure 3. 2) or continuous (Figure 3. 3).

We shall first consider how to determine the probability distribution of a discrete random variable.

The number of times each voltage step in Figure 3. 2 occurs is counted:

$$u_1 - n_1 \text{ times},$$

$$u_2 - n_2 \text{ times},$$

$$u_k - n_k \text{ times}.$$

These numbers designated n_1 to n_k are divided by N, the number of measurements made during the particular time interval. This gives the distribution sequence in Table 3.1.

TABLE 3.1

Values of discrete voltage	u_1	u_2	u_3	u_4	...	u_k
Recurrence of each value	n_1	n_2	n_3	n_4	...	n_k
Probability of each value	$P_1=\frac{n_1}{N}$	$P_2=\frac{n_2}{N}$	$P_3=\frac{n_3}{N}$	$P_4=\frac{n_4}{N}$	...	$P_k=\frac{n_k}{N}$

In addition, the probability distribution function of a discrete random variable can be determined. This function can be constructed from the parameters u_1, u_2, u_k, and N. The discrete values of a voltage from Figure 3.2 are marked off on the abscissa in Figure 3.4, and on the ordinate the ratio of the number of values, at which the voltage is lower than the value specified, to the total number of measurements (N).

For example, for $u_2 \geqslant u \geqslant u_1$, the ordinate is $\frac{n_1}{N}$, for $u_3 \geqslant u \geqslant u_2$ the ordinate is $\frac{n_1+n_2}{N}$, for $u_4 \geqslant u \geqslant u_3$ — $\frac{n_1+n_2+n_3}{N}$, etc.

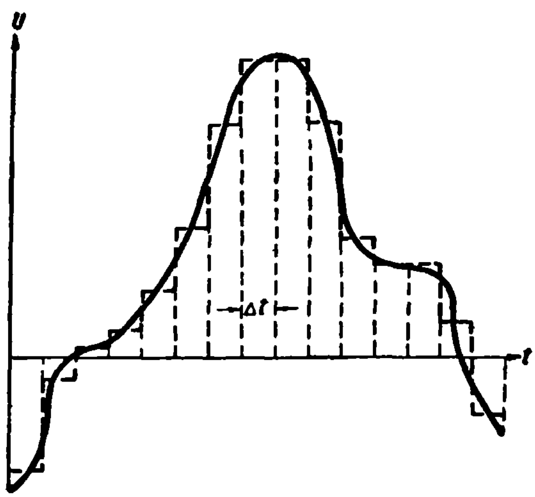

FIGURE 3.3. Continuous random function.

The resulting curve is the probability distribution function of a discrete random variable.

The probability distribution function of a continuous random variable is derived by replacing the continuous random function (Figure 3.3) with a stepwise curve at constant intervals Δt. The distribution function is plotted

as previously, the only difference being that the resulting discrete probability distribution function is approximated as far as possible by a continuous curve (Figure 3.5).

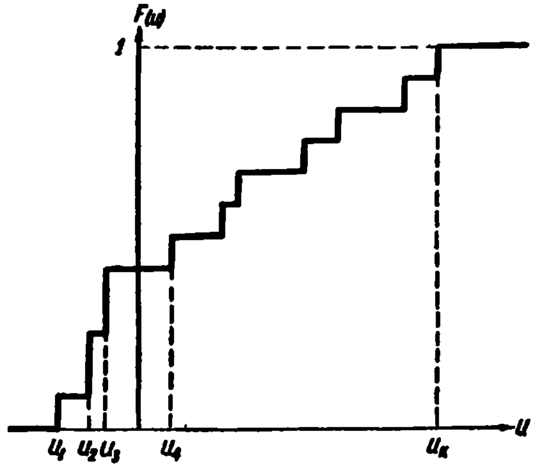

FIGURE 3.4. Probability distribution function of a discrete variate.

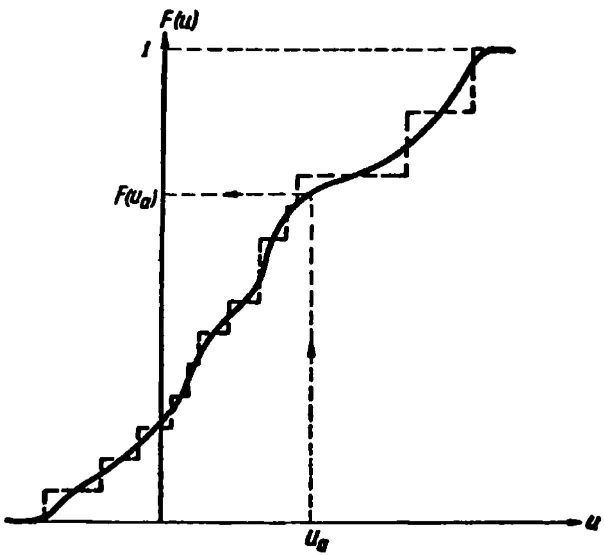

FIGURE 3.5. Probability distribution function of a continuous random variable.

Given the probability distribution function of either a discrete or a continuous random variable, the probability that any voltage specified by the

random function is less than a preassigned voltage u_a (Figure 3.5) may be determined.

Another curve describing a continuous random variable is the differential probability distribution function.

The curve (see Figure 3.6) is obtained by differentiating the probability distribution function:

$$p(u) = \frac{dF(u)}{dt}. \tag{3.8}$$

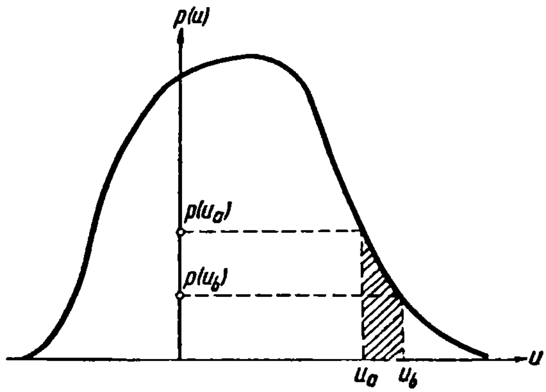

FIGURE 3.6. Differential probability distribution function (probability density distribution) of a continuous variate.

To determine the differential distribution function from a continuous random function of time, the latter should again be replaced by a step curve. In practice, the procedure is simplified as follows (Figure 3.7). We decide on a time interval Δt and draw at these constant intervals (see the bottom graph) straight lines meeting the random function at N points. Lines parallel to the time axis are then drawn at arbitrary constant intervals Δu. Dividing the number of points within each Δu interval by the total number of points (N), we obtain the corresponding ordinate of the differential function (the dashed lines in the upper curve in Figure 3.7). The resulting smoothened step curve is the differential distribution function of the continuous random variable.

We shall now examine a distinctive feature of the differential distribution function of a continuous random variable (Figure 3.6) which is not apparent from its distribution sequence.

Each element of the distribution sequence gives the probability that any randomly measured voltage will be equal to a given value. For example, P_1 gives the probability that randomly measured voltage is equal to u_1 (Table 3.1).

Suppose now that the ordinates of the dashed curve in the upper graph of Figure 3.7 also give the probability that measured voltage will fall in a given interval Δu. The probability that the measured voltage falls in one of

the k given Δu intervals is equal to one:

$$P_1+P_2+\ldots+P_k=1, \tag{3.9}$$

where P_1 = the probability that the measured voltage falls in the first Δu interval;

P_2 = the probability that the measured voltage falls in the second Δu interval, etc.

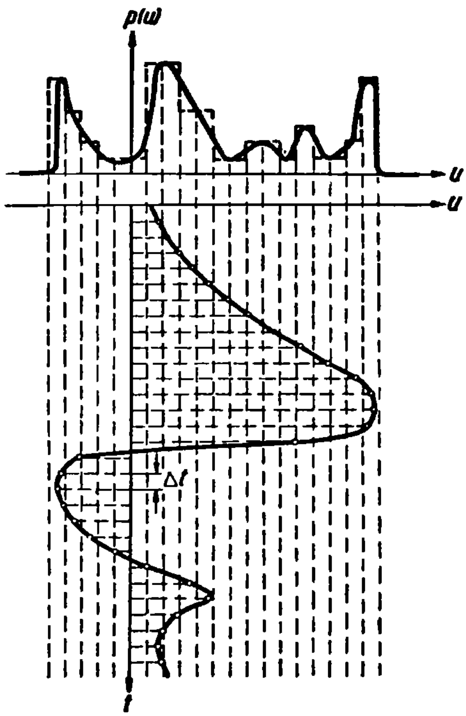

FIGURE 3.7. Construction of the differential probability distribution function of a continuous random variable.

If the number k of Δu intervals is increased, we readily see that as $k\to\infty$, each of these intervals is reduced to a point. In this case the probability that any randomly measured voltage falls in one of the intervals reduces to the probability that the measured voltage is equal to a given value (the interval being reduced to a point).

On the other hand, the sum of these k probabilities remains constant and equal to 1 (see 3.9). Therefore (at $k=\infty$) the probability that the measured voltage falls at a given point (i.e., is equal to a preassigned value) is zero.

Neverthless, none of the ordinates of the differential distribution function plotted in Figure 3.7 for a continuous random variable is zero, because,

when passing from a discrete to a continuous random variable, the ordinate no longer gives the probability (P) but rather the probability density (p), defined as the ratio of the probability to the voltage interval in question:

$$p=\frac{P}{du},$$

where du is the width of the interval at $k=\infty$.

The differential probability distribution function of a continuous variate is therefore also called the probability density distribution.

The probability that a randomly measured voltage falls between u_a and u_b (Figure 3.6) is determined from the density distribution by integration:

$$P_{a-b}=\int_{u_a}^{u_b} p\, du. \tag{3.10}$$

The probability of all possible events is 1 (3.4), therefore

$$P=\int_{-\infty}^{\infty} p\, du=1. \tag{3.11}$$

In other words, the area between the curve $p(u)=f(u)$ and the abscissa (u) is always normalized to 1. The probability that a given voltage falls between $u_b \geqslant u \geqslant u_a$ is graphically determined (Figure 3.6) as the ratio of the hatched area to the total area between the curve $p(u)=f(u)$ and the abscissa.

When dealing with random functions, we should remember that there exist two different concepts for the average values of these functions.

First, there is the time average of a random variable. For a discrete random function (Figure 3.2) the time average of the random variable is defined as

$$u_{\text{av}}=\frac{u_1+u_2+\ldots+u_k}{N}=\frac{\sum\limits_{m=1}^{k} u_m}{N},$$

where u_1—u_k = the voltages on the corresponding Δt portions;
N = the total number of Δt intervals constituting the range of the random function (T).

Multiplying the numerator and the denominator by Δt, we obtain

$$u_{\text{av}}=\frac{1}{T}\sum_{m=1}^{k} u_m\, \Delta t, \tag{3.12}$$

where $T=N\,\Delta t$ is the range of the random function.

For a continuous random function $(\Delta t \to 0)$, (3.12) takes on the following form:

$$u_{av} = \frac{1}{T}\int_0^T u\,dt.$$

If the function must be averaged over an infinite time interval (from $-\infty$ to $+\infty$), this equation is slightly modified:

$$u_{av} = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} u\,dt. \tag{3.13}$$

This equation is the familiar one used to determine the average of both random and determinative variables.

In addition, random variables can be characterized by their statistical average, also called mathematical expectation. This average is defined for a discrete random function by the equation

$$u_0 = \sum_{m=1}^{k} P_m u_m, \tag{3.14}$$

where u_m = the voltage of the m-th interval Δt (Figure 3.2);
P_m = the probability of this voltage.

This probability can be expressed in terms of the probability density:

$$P_m = p_m \Delta u.$$

Equation (3.14) can therefore be somewhat modified:

$$u_0 = \sum_{m=1}^{k} p_m u_m \Delta u.$$

For a continuous random function $(\Delta u \to 0)$ we therefore have

$$u_0 = \int_{-\infty}^{\infty} pu\,du. \tag{3.15}$$

The abscissa of the center of gravity of the plane homogeneous figure (3.6) enclosed between the function $p=f(u)$ and the (u) axis is given by

$$u_c = \frac{\int_{-\infty}^{\infty} pu\,du}{S},$$

where S is the area of the figure. But, according to (3.11), this area S is always equal to 1. Therefore, (3.15) determines the abscissa of the center of gravity of this figure.

We have previously observed that under certain conditions all stationary random processes have the so-called ergodic property. It follows from this

property (we omit the proof of the proposition) that for ergodic processes the statistical average of a random variable is always equal to its time average. We may therefore write

$$u_{av} = u_0. \tag{3.16}$$

The deviation of a random variable is defined as the difference

$$\Delta u = u - u_0, \tag{3.17}$$

where u = the current value of the random variable;

u_0 = the statistical average (mathematical expectation).

The theory of probability often makes use of the variance of a random variable. Variance characterizes the intensity of a random variable about its average, and is defined by the equation

$$D = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} (u - u_0)^2 p \, dt, \tag{3.18}$$

where p is the probability density distribution. If the variance is zero, the random variable is reduced to a constant value equal to u_0.

To obtain a parameter whose dimensions are those of the random variable (u) and of its statistical average (u_0), the variance is often replaced by the root-mean-square or standard deviation of the random variable:

$$\sigma = \sqrt{D}. \tag{3.19}$$

We have shown previously that the probability density distribution can be determined from experimental data. However, the probability density of fairly many random variables have normal (or Gaussian) distribution. This distribution is defined as

$$p = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(u-u_0)^2}{2\sigma^2}}, \tag{3.20}$$

where σ = the root-mean-square (rms) deviation;

u_0 = the statistical average of the random variable.

The normal probability density distribution is represented by the curve shown in Figure 3.8.

The normal distribution defined by (3.20) is always symmetrical about the vertical axis passing through its maximum. The statistical average u_0, (see (3.15)) gives the abscissa of the center of gravity of Figure 3.8 and determines in this case the voltage at which the maximum of the normal distribution is attained:

$$\frac{1}{\sqrt{2\pi}\,\sigma}.$$

If a random function of time has no d-c component, the average (3.16) is zero. In this case (Figure 3.9) the normal-distribution curve is symmetrical about the ordinate axis.

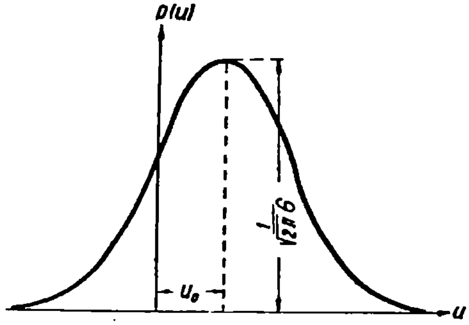

FIGURE 3.8. Normal probability density distribution.

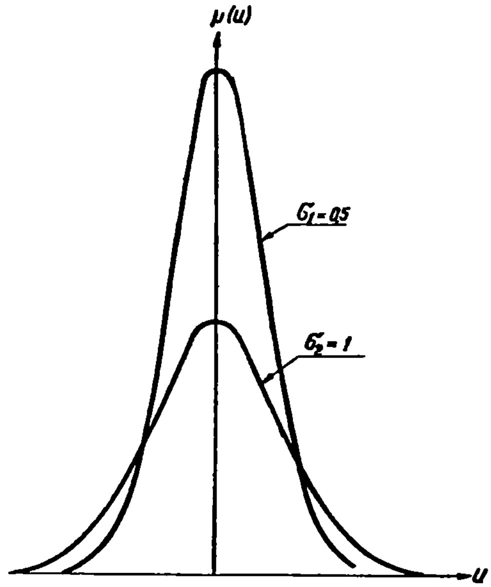

FIGURE 3.9. Normal distribution of a random variable with no d-c component.

It follows from (3.20) that the rms deviation σ is a measure of the steepness of the normal distribution curve. The effect of σ on the shape of this curve is illustrated in Figure 3.9.

3. ENTROPY OF A RANDOM EVENT

Electrical current is used for two purposes. First, as a carrier of energy to various systems and assemblies; second, to transmit information.

When electrical current functions as an energy-transferring medium, we are mostly concerned with the efficiency of transfer, since it is desirable to transfer energy with minimum losses.

On the other hand, a distinctive feature of any transmission of information is its independence of the amount of energy lost in transmitting the current which carries this information. In other words, an equal amount of information can be transmitted by currents of different amplitudes. When current is used as an information-transmitting medium, the problems involved in maximizing the efficiency are therefore of no primary importance. The main consideration is maximizing the amount of information transmitted in a given time interval.

When dealing with physical systems (automatic-control systems included), we know some relevant facts, but there are some parameters and characteristics which are unknown to us. In this case we say that the system is indeterminate. A measure of this indeterminacy is the entropy H.

Consider the following experiment. Take a system whose entropy is H_{ini}. When we have received some information on this system, its entropy decreases to H_{fin}.

The amount of information (q) received is defined as

$$q = H_{\text{ini}} - H_{\text{fin}}. \tag{3.21}$$

The amount of information is thus equal to the entropy decrement produced by the message received. When the message received makes the physical system fully determinate ($H_{\text{fin}} = 0$), the amount of information is numerically equal to the initial entropy of the system:

$$q = H_{\text{ini}}. \tag{3.21a}$$

a) Determination of entropy of a discrete signal

In the simplest case the entropy is due to the possibility of receiving either of two messages: "the event has occurred" or "the event has not occurred". This can be written briefly as

"YES-NO" or "1-0".

The unit of entropy is therefore defined as the uncertainty involved in the choice of one of the two possible outcomes YES-NO or 1-0 (the binary unit). Since the amount of information (3.21) is measured in the same units, this signal also defines the unit of information.

We shall now determine the entropy of a simple event when we are faced with a probable choice of one out of m possible outcomes.

Consider the following example. Find the entropy of selecting one out of eight cells (Figure 3.10).

It follows from Figure 3.10 that, in order to state which cells have been selected, we must know the positions of three switches.

These data are arranged in Table 3.2.

TABLE 3.2

Cell number	0	1	2	3	4	5	6	7
Switch I	0	0	0	0	1	1	1	1
Switch II	0	0	1	1	0	0	1	1
Switch III	0	1	0	1	0	1	0	1

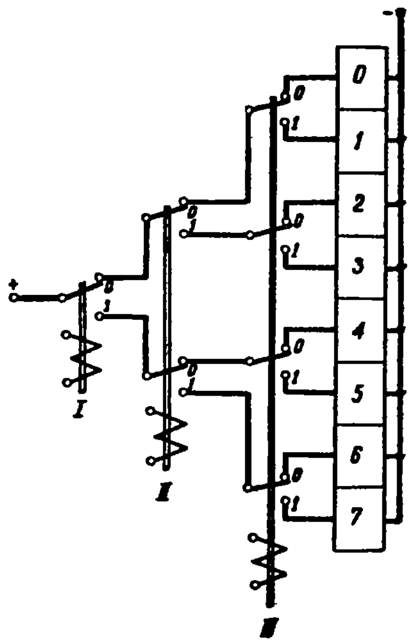

FIGURE 3.10. Selection circuit for one out of eight cells.

The entropy involved in the selection of one out of eight cells is thus three (the position of three switches is uncertain). Mathematically this is stated as

$$(H)_8 = 3.$$

This expression can be slightly modified:

$$(H)_{2^3} = 3.$$

The entropy is thus equal not to the number of cells (here 8), but to the power to which the base 2 is raised to give the number of cells ($2^3 = 8$ so that $H = 3$).

Therefore, in general, the entropy of an event (the uncertainty involved in the occurrence of one out of m possible outcomes) is equal to

$$H = \log_2 m. \tag{3.22}$$

This definition of the entropy of a simple event is known as Hartley's law.

In the previous example $m=8$; the entropy involved in the choice of one cell out of eight is therefore equal to

$$H = \log_2 m = \log_2 8 = 3.$$

Equation (3.22) is also assumed to apply when m is not an integer.

Information theory is very closely associated with the theory of probability. Indeed, let all the m values of the random variable considered in the preceding be equiprobable. The probability that one of them occurs is therefor equal to

$$P = \frac{1}{m}.$$

In this case (3.22) the entropy of the random event entailing the occurrence of one of the m values is written in the form

$$H = \log_2 \frac{1}{P}. \tag{3.23}$$

We thus see that the higher the probability of an event, the lower its entropy (uncertainty). Indeed, if we know with certainty (with probability $P=1$) that cell No. 7 in Figure 3.10 will be selected, the entropy of the event (the uncertainty of cell No. 7 being selected) is zero.

Equation (3.23) can be written in a slightly different form:

$$H = \log_2 \frac{1}{P} = \log_2 1 - \log_2 P.$$

But

$$\log_2 1 = 0..$$

Therefore,

$$H = -\log_2 P. \tag{3.24}$$

This equation determines the entropy of a single (simple) event. A compound event consisting of n equiprobable simple independent events has n times as much entropy (3.6a):

$$H = -n \log_2 P. \tag{3.25}$$

It should be kept in mind that the probability (3.2) is always between the limits $1 \geqslant P \geqslant 0$. Therefore, $\log_2 P$ is negative, i.e., the entropy is always positive.

Consider the case when different events have different probabilities. By analogy with equation (3.24), we write for each event

$$H_k = -\log_2 P_k, \tag{3.26}$$

where k is the index of the event ($k = 1, 2, \ldots, m$).

Equation (3.26) determines the entropy of a single event whose probability is P_k.

Further consider m different events with the following probabilities:

$$\text{event } A_1 — P_1;$$
$$\text{event } A_2 — P_2;$$
$$\cdots\cdots\cdots$$
$$\text{event } A_m — P_m.$$

In general, several events may occur simultaneously. The events in question are therefore joint. They are, moreover, independent.

If event A_1 has the probability P_1, then the probability (3.6a) of two events A_1 occurring simultaneously is equal to

$$P(2A_1) = P_1 P_1$$

or

$$P(2A_1) = P_1^2.$$

The probability that event A_1 occurs a times is naturally

$$P(aA_1) = P_1^a. \tag{3.27}$$

Further consider a compound event consisting of the simultaneous occurrence of n simple events $A_1, A_2, \ldots, A_m$.

The probability that event A_1 recurs nP_1 times in the compound event (in n simple events) is from equation (3.27) equal to

$$P[(nP_1)A_1] = P_1^{nP_1}.$$

Similarly, the probability that event A_2 recurs nP_2 times is equal to

$$P[(nP_2)A_2] = P_2^{nP_2}.$$

Therefore, from equation (3.6a), the probability of the compound event consisting of simple events $A_1, A_2, \ldots, A_m$ is equal to

$$P = P_1^{nP_1} P_2^{nP_2} \ldots P_m^{nP_m}.$$

Hence, from equation (3.24), the entropy of the compound event is determined by

$$H = -\log_2 (P_1^{nP_1} P_2^{nP_2} \ldots P_m^{nP_m}).$$

Further manipulation with this equation gives

$$H=-nP_1\log_2 P_1-nP_2\log_2 P_2-\ldots-nP_m\log_2 P_m$$

or

$$H=-n\sum_{k=1}^{m}P_k\log_2 P_k. \tag{3.28}$$

Equation (3.28), called Shannon's equation, gives the entropy of a compound event consisting of simultaneous occurrence of n simple events of m kinds whose probabilities are different.

If the events of all m kinds are equiprobable ($P_1=P_2=\ldots=P_m=P$), Shannon's equation reduces to

$$H=-nP\sum_{k=1}^{m}\log_2 P=-nmP\log_2 P.$$

Since the probability of m equiprobable events is $P=\frac{1}{m}$, the preceding equation can be rewritten as

$$H=-n\log_2 P.$$

We thus see that equation (3.25), as could have been expected intuitively, is a particular case of Shannon's equation.

When a simple event ($n=1$) is considered, equation (3.28) is written as

$$H=-\sum_{k=1}^{m}P_k\log_2 P_k. \tag{3.28a}$$

The reader should observe one essential difference between equations (3.24) and (3.28a). Equation (3.24) gives the entropy of one out of m equiprobable events. Equation (3.28a), on the other hand, gives the average entropy for one out of m different events whose probabilities are different.

It is noteworthy that according to the second law of thermodynamics (Boltzmann's law), the entropy of closed space is

$$H=-\frac{1}{N}\sum n_k\ln\frac{n_k}{N}, \tag{3.29}$$

where N = the total number of molecules in the given space;
n_1 = the number of molecules having the velocity $V_1+\Delta V$;
n_2 = the number of molecules having the velocity $V_2+\Delta V$, etc;
ΔV = the velocity increment.

We have previously shown in equation (1.4) that

$$\ln\frac{n_k}{N}=n\log_2\frac{n_k}{N},$$

where n is a coefficient .

The ratio $\frac{n_k}{N}$ is in fact the probability that the molecules have the velocity $V_k+\Delta V$. We may therefore put

$$P_k=\frac{n_k}{N}.$$

The entropy (3.29) is thus determined by the expression

$$H=-n\,\Sigma P_k \log_2 P_k. \tag{3.30}$$

Comparing this equation with Shannon's law (3.28), we see why Shannon chose the word entropy to describe the degree of uncertainty of a physical system.

Originally the concept of entropy was used only in thermodynamics, where entropy characterized the irreversible dissipation of energy in a closed physical system and thus could only increase. In cybernetics, however, entropy may either increase or decrease, since this concept is closely connected with the measure of uncertainty. An increase in the amount of *information will lower the uncertainty, i.e., the entropy will decrease.*

There is no contradiction between the statement that, in cybernetics, entropy may equally well increase and decrease, whereas in physical systems entropy may only increase. This is so, because thermodynamics is concerned with <u>limited</u> systems, whereas the universe is <u>infinite</u>, and there are apparently areas where new "solar" systems are being created and the local (thermodynamic) entropy decreases.

Let us consider an example of entropy determination.

In Riga, trolleybuses Nos 4, 6, and 7 travel down Lenin Street. Trolleybus No. 4 passes every 3 minutes, No. 6 every 4 minutes, and No. 7 every 5 minutes. Determine the average entropy of communications on all three trolleybus lines.

The probability that any of the trolleybuses arrives is 1. Therefore,

$$P_4+P_6+P_7=1, \tag{3.31}$$

where P_4, P_6, P_7 are, respectively, the probabilities that trolleybuses No. 4, No. 6 and No. 7 arrive.

Moreover, given the data on the frequency of the trolleybuses, we have

$$3P_4=4P_6,$$
$$3P_4=5P_7.$$

Inserting P_6 and P_7 into (3.31), we find the probability of trolleybus No. 4 arriving:

$$P_4=\frac{20}{47}=0.426.$$

In addition, from (3.31) we obtain the probabilities of trolleybuses No. 6 and No. 7 arriving:

$$P_6=0.319;$$
$$P_7=0.255.$$

Given the probability of each event, we find from (3.28a) the average entropy of these events:

$$H=-P_4\log_2 P_4-P_6\log_2 P_6-P_7\log_2 P_7=$$
$$=-0.426\log_2 0.426-0.319\log_2 0.319-0.255\log_2 0.255=1.55.$$

In a binary-number system there are two random events (an outcome of 1 and 0). We shall now consider a compound random event consisting of two ($m=2$) simple random events.

The average entropy for each of these simple events is from (3.28a):

$$H=-P_1\log_2 P_1-P_0\log_2 P_0, \tag{3.32}$$

where P_1 and P_0 are the probabilities that the outcome is 1 and 0.

Let the probability of an outcome of 1 be

$$P_1=P.$$

Since the probability of all (in this case, two) events is equal to one,

$$P_1+P_0=1,$$

the probability of the second event is given by

$$P_0=1-P.$$

Inserting the probabilities P_1 and P_0 into (3.32), we obtain the average entropy per simple event (the outcome of 1 or 0):

$$H=-P\log_2 P-(1-P)\log_2(1-P). \tag{3.33}$$

This relationship is plotted in Figure 3.11. Differentiating (3.33) we easily find that the entropy is maximum when the two events are equiprobable:

$$P=P_1=P_0=\frac{1}{2}. \tag{3.34}$$

When binary-coded information is transmitted, the amount of information delivered by a single signal is highest when the entropy is maximum (see (3.21a)). We should therefore try to dispatch messages coded so that ones and zeros are as close to equiprobable as possible.

Until now we have only considered independent events. As regards dependent events, the following should be observed.

The dependence of the probability of event B on the outcome of event A in itself gives some information on event B, thus reducing the entropy of this event. The entropy of dependent events (H_{con}) is therefore always less than the entropy (H) of independent events:

$$H_{\text{con}}<H. \tag{3.35}$$

It is generally said that c o r r e l a t i o n (dependence) between events lowers the entropy (uncertainty).

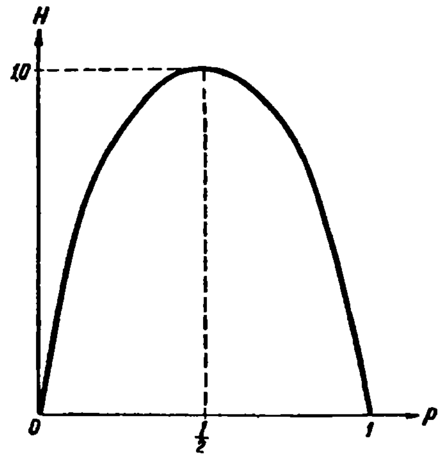

FIGURE 3. 11. Average entropy as a function of the probability of one binary event.

b) Measuring the entropy of a continuous signal

We have previously shown in section 1-5 that a continuous function of time can be represented with any desired accuracy by a finite number of its values. In other words, given the admissible error, we may determine the minimum increment Δu allowed, thus replacing the continuous function by a discrete curve (analog-to-digital conversion).

In practice, this minimum increment is determined from the equation

$$\Delta u = \frac{2\varepsilon}{100}(u_{max} - u_{min}), \tag{3.36}$$

where ε = the admissible error in percent; this error is multiplied by 2, since the error specified refers to deviations from either side of the true value;

u_{max} = the maximum value of the continuous function in the given range;

u_{min} = the minimum value of the function in this range.

The continuous curve $u = f(t)$ is divided into m steps of Δu width. The original curve is now replaced by a stepped curve (Figure 3.12) consisting of m discrete (digital) values:

$$m = \frac{u_{max} - u_{min}}{\Delta u}. \tag{3.37}$$

If now equations (3.36) and (3.37) are solved simultaneously, we obtain

$$m=\frac{50}{\varepsilon}. \tag{3.38}$$

An analysis of a continuous random variable is thus reduced to the analysis of a discrete random variable assuming m values. When all m events are equiprobable, we can easily determine the probability of each event using (3.38):

$$P=\frac{1}{m}=\frac{\varepsilon}{50}. \tag{3.39}$$

The entropy of an event (the entropy of the occurrence of any step) is determined from (3.24):

$$H=-\log_2 P=-\log_2\left(\frac{\varepsilon}{50}\right)=\log_2\left(\frac{50}{\varepsilon}\right). \tag{3.40}$$

If the different events (occurrence of different steps) have different probabilities, their entropy is determined by Shannon's equation (3.28a).

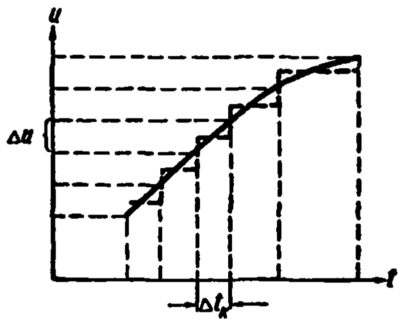

FIGURE 3.12. Substitution of a discrete curve for a continuous curve.

The minimum increment Δu is still calculated from (3.36) and is a constant for the entire range. The derivative of the continuous curve, however, is not constant over this range. Therefore, the width (Δt_k) of each step in Figure 3.12 is also not constant.

Let us now determine the minimum time interval (Δt_{min}).

$$\left(\frac{du}{dt}\right)_{max}=u'_{max}=\frac{\Delta u}{\Delta t_{min}}.$$

Hence

$$\Delta t_{min}=\frac{\Delta u}{u'_{max}}, \tag{3.41}$$

where Δu = the increment given by (3.36);

u'_{max} = the maximum rate of change of the continuous curve in the given time interval.

Inserting Δu from (3.36) into (3.41), gives

$$\Delta t_{min} = \frac{\varepsilon}{50 u'_{max}} (u_{max} - u_{min}). \tag{3.42}$$

Dividing the entropy of an event (the occurrence of a step) by the duration (in time) of the corresponding step, gives the frequency with which the information transmitted by the particular step is received:

$$f_q = \frac{H}{\Delta t}. \tag{3.43}$$

For the minimum time interval we have

$$f_{q_{max}} = \frac{H}{\Delta t_{min}}. \tag{3.44}$$

Inserting H from (3.40) and Δt_{min} from (3.42), we obtain

$$f_{q_{max}} = \frac{50}{\varepsilon} \cdot \frac{u'_{max}}{u_{max} - u_{min}} \log_2 \left(\frac{50}{\varepsilon}\right). \tag{3.45}$$

Equation (3.45) gives the maximum frequency with which a receiver should receive the information transmitted by a continuous signal.

Example. Find the maximum frequency with which the information transmitted by a sinusoidal curve ($f = 50$ cycles per second) varying from zero to maximum ($u_{max} = 10$ v)(Figure 3.13) should be received; admissible error $\varepsilon = 5\%$.

We first determine the step of the curve approximating the sinusoidal function. From (3.36) we have

$$\Delta u' = \frac{2\varepsilon}{100}(u_{max} - u_{min}) = \frac{2 \cdot 5}{100}(10 - 0) = 1 \text{ v}.$$

The number of steps is equal from (3.38) to

$$m = \frac{50}{\varepsilon} = \frac{50}{5} = 10.$$

The sinusoidal curve is thus approximated by a ten-step curve.

We further find the probability of each of the ten events in question (of each of the ten voltage steps). A sinusoidal function is a periodic curve where the probabilities of the respective voltage values (disregarding the sign) recur every quarter cycle. To determine all the probabilities, it therefore suffices to consider the sinusoidal curve for $\frac{\pi}{2} > \theta > 0$.

The ordinate in Figure 3.13 is divided into steps of $\Delta \upsilon = \frac{\pi}{72}$. This gives 37 points (including the origin) on the sinusoidal curve. Now count the points

in each of the ten ($m=10$) voltage intervals (a point occurring at the boundary of two intervals will be counted as $\frac{1}{2}$ in each interval) and divide the count by the total number of points (37). We now determine the probability of each of the ten voltage steps (the random events). The data are arranged in Table 3.3.

TABLE 3.3

Number of step	1	2	3	4	5	6	7	8	9	10
Voltage interval, v	0–1	1–2	2–3	3–4	4–5	5–6	6–7	7–8	8–9	9–10
Probability	0.0675	0.0675	0.0675	0.0675	0.0675	0.0675	0.081	0.095	0.122	0.297

From these data we plot the step curve shown in the upper graph in Figure 3.13.

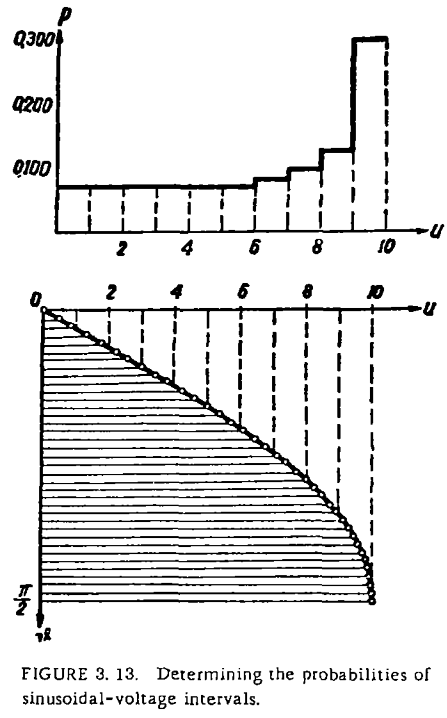

FIGURE 3.13. Determining the probabilities of sinusoidal-voltage intervals.

Since the probabilities of the individual voltage steps are different, the average entropy of one out of ten ($m=10$) events is found from Shannon's equation (3.28a):

$$H=-\sum_{k=1}^{m} P_k \log_2 P_k=-\sum_{k=1}^{10} P_k \log_2 P_k.$$

Inserting the probabilities from Table 3.3, we obtain

$$H=-(6\cdot 0.0675\log_2 0.0675+0.081\log_2 0.081+0.095\times$$
$$\times\log_2 0.095+0.122\log_2 0.122+0.297\log_2 0.297)=3.07.$$

If now all the steps are taken as equiprobable (in practice this means that the sinusoidal curve is replaced by a straight line in the $\frac{\pi}{2}\geqslant\theta\geqslant 0$ interval), the entropy of each event is determined from (3.40):

$$H=\log_2\left(\frac{50}{\varepsilon}\right)=\log_2\left(\frac{50}{5}\right)=3.33.$$

The calculations are naturally simplified, but the result is less accurate.

To determine the maximum frequency of information in equation (3.45) we must know the maximum rate of change of the voltage. This rate of change is very simple to determine.

The sine function is analytically defined as

$$u=u_{max}\sin 2\pi ft,$$

therefore, by differentiation,

$$\frac{du}{dt}=2\pi f u_{max}\cos 2\pi ft.$$

This derivative has its maximum at $2\pi ft=0$, since $\cos 0=1$ 1. Hence,

$$\left(\frac{du}{dt}\right)_{max}=u'_{max}=2\pi f u_{max}=2\pi 50\ 10=3140 \text{ v/sec}.$$

The maximum frequency for which the receiver should be adapted is from (3.45) equal to

$$f_{q_{max}}=\frac{50}{\varepsilon}\cdot\frac{u'_{max}}{u_{max}-u_{min}}\log_2\left(\frac{50}{\varepsilon}\right)=$$

$$=\frac{50}{5}\cdot\frac{3140}{10-0}\log_2\left(\frac{50}{5}\right)=10\,500 \text{ binary units/sec}.$$

4. OPTIMUM CODE

Let us now consider the coding of information transmitted by a discrete binary signal (1-0).

We have previously shown in section 3-3a that the entropy of a binary signal is maximum when two conditions are satisfied:

a) 1 and 0 are equiprobable:

$$P_0 = P_1 = \frac{1}{2};$$

b) all the signals are independent.

Suppose that binary information is transmitted as follows: 1 is constant voltage of duration τ, and 0 is the absence of voltage during τ. The number of signals (pulses) transmitted per second through the channel (i.e., the communication link) is therefore

$$n = \frac{1}{\tau}.$$

The channel capacity (the maximum number of signals which can be transmitted through the channel in one second) is therefore

$$C = \frac{1}{\tau}. \tag{3.46}$$

A code will be called optimum if it ensures the transmission of the maximum amount of information through the channel. Shannon proved that, regardless of the actual optimum code, the average rate of transmission of information through the channel cannot exceed

$$V_1 = \frac{C}{H}, \tag{3.47}$$

where C = the channel capacity;

H = the entropy of the message transmitted.

This equation follows from Shannon's first theorem.

Consider an example for the determination of the maximum rate of transmission (V_1).

The Latvian alphabet consists of 30 letters ($m=30$), including the "blank", i.e., space between two words. For the sake of simplicity, let all the letters be equiprobable. The entropy of a single letter from (3.24) is then equal to

$$H = -\log_2 P.$$

Since all the letters are assumed equiprobable, we have for each letter

$$P = \frac{1}{m} = \frac{1}{30} = 0.0333.$$

The entropy of a single letter is therefore

$$H = -\log_2 0.0333 = 4.93.$$

Let the channel capacity be C = 1000 pulses per second. Regardless of the actual method of binary coding, the contents of the message cannot be

transmitted through the channel faster than that given by (3.47):

$$V_1 = \frac{C}{H} = \frac{1000}{4.93} = 203 \text{ letters/sec.}$$

To determine the optimum code, we must take into consideration some statistical data on the message transmitted. For the most frequently occurring signals the code should be as short as possible, whereas the infrequent signals can be coded by a longer succession of bits.

A practical optimum code, the Shannon-Fano code, is constructed as follows.

All the signals to be coded are divided into two groups so that the probability of each group be as close to $\frac{1}{2}$ as possible. This maximizes the information (the entropy of the message is at its maximum). One of the two groups is then coded as 1 and the other as 0. The signals collected in each group are again divided into two groups with probabilities close to $\frac{1}{2}$. Each of the secondary groups is again coded as 1 and 0, and are placed in the second position. This process is repeated until each group comprises one signal only.

Let us apply the Shannon-Fano procedure to the following example.

Consider 10 signals whose probabilities are given in Table 3.4.

The code is compiled in five stages, as we see from the table. In the first stage the signals are divided into two approximately equiprobable groups. In the first group (a_1) we have signals Nos 0 and 1, whose total probability is 0.45, and in the second group (b_1) signals Nos 2 through 9 whose total probability is 0.55. The signals of the first group are coded as 1, and those of the second group as 0.

TABLE 3.4

Number of signal	Probability of signal	Stages I	Stages II	Stages III	Stages IV	Stages V	Code I	Code II	Code III	Code IV	Code V
0	0.25	a_1	a				1	1			
1	0.20		b_2				1	0			
2	0.15	b_1	c_2	a_3			0	1	1		
3	0.10			b_3			0	1	0		
4	0.10		d_2	c_3	a_4		0	0	1	1	
5	0.05				b_4		0	0	1	0	
6	0.05			d_3	c_4	a_5	0	0	0	1	1
7	0.05					b_5	0	0	0	1	0
8	0.03				d_4	c_5	0	0	0	0	1
9	0.02					d_5	0	0	0	0	0

When the most significant bit is 1, the code refers to the first group of signals, and when it is 0 the code refers to the second group.

In the second stage the first group of signals (a_1) is divided into two and signal No. 0 is coded by 1 (in the next most significant position), while signal No. 1 is coded by 0. The second group of signals (b_1) is also divided into two approximately equiprobable parts. The first part (c_2) is coded by 1 in

the next most significant position (column III), and the second group (d_2) by zero. The process is continued until finally each group contains but one signal.

Note that ten signals can be coded (see (1.3)) by four-bit numbers. In the optimum code, however, the different signals have from two to five bits. The higher the probability of a signal, the shorter its code.

The optimum Shannon-Fano code is therefore constructed so that the component signals (1 and 0) are equiprobable as far as possible.

Until now we have considered only noiseless channels.

Let us now consider the effect of noise in the previous example. If the channel is noiseless, then given the code we can identify each signal received with a probability of one. Suppose that we have thus identified signal No. 3 with full certainty (Figure 3.14,a)

If the signal is accompanied by noise, we cannot identify it with a probability of one, saying, e.g., that it is certainly No. 3 (Figure 3.14, b). In this case all we can say is that most probably this is signal No. 3, but the possibility is not excluded of this being a different signal. The probability distribution of a message concerning a signal is shown in Figure 3.14, b.

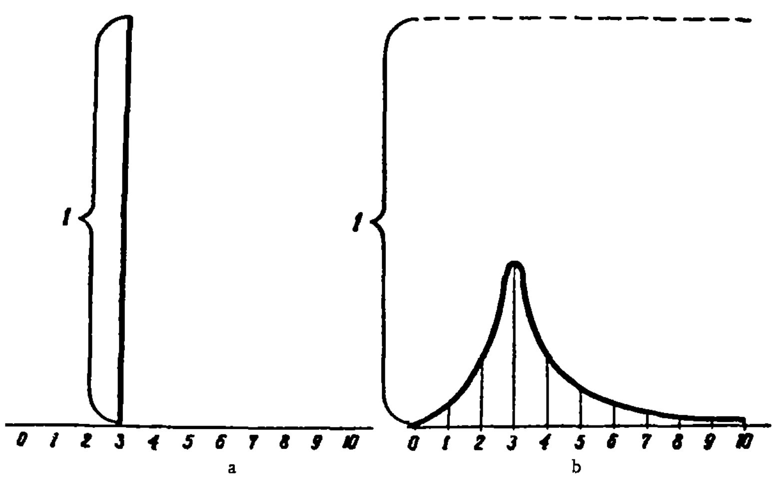

FIGURE 3.14. Probability when a message is received.

If there is noise some information is lost. This loss is determined by the equation

$$\Delta H = H_{\text{in}} - H, \tag{3.48}$$

where H_{in} = the amount of information carried by the signal at the channel input;

H = the amount of information in the signal at the channel output.

ΔH characterizes the average measure of uncertainty remaining after the signal is received.

The maximum rate of message transmission in a noiseless channel is determined by (3.47). If there is noise, this rate is reduced owing to loss

of information as given by (3. 48). Therefore,

$$V_2 = \frac{C}{H_{\text{in}}} = \frac{C}{H + \Delta H}. \tag{3.49}$$

This relationship follows from Shannon's second theorem. In real channels the velocity V_1, and even more so V_2 (3. 47), cannot be attained, and we still do not know what improvements are necessary to attain these rates of information transmission.

Chapter IV

STATISTICAL METHODS FOR THE ANALYSIS OF AUTOMATIC-CONTROL SYSTEMS

Automatic-control systems are generally designed assuming a sinusoidal, step, or pulse input signal. The complexity of cybernetic systems and the great variety of their operating conditions require methods of design for ensuring their response to arbitrary signals whose shape is not known in advance. These systems and their behavior are therefore studied and analyzed using probability theory and statistical methods.

As a measure of error in systems studied by statistical methods we often take the mean square error, defined by the equation

$$e^2 = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} [u_d(t) - u_2(t)]^2 dt, \tag{4.1}$$

where $u_d(t)$ = the desired (continuous) output signal;
$u_2(t)$ = the actual (continuous) output signal;
$e = u_d(t) - u_2(t)$ = the output error of the system;
T = the relevant period.

Another parameter characterizing the operation of the system is the simple mean error

$$e_m = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} [u_d(t) - u_2(t)]\, dt. \tag{4.2}$$

When the mean error is used as a measure of the system operation, individual errors introduced by the different values of the random variable are averaged according to their magnitude.

Introduction of the concept of mean square error leads to the following: in a system optimized relative to e^2 the large errors are reduced at the expense of a slight increase of the small errors. In other words, in this case the larger the error, the greater its significance in calculations.

The object of statistical analysis of cybernetic automatic-control systems is to minimize the mean square (in most cases) or some other error.

1. TRANSFER FUNCTIONS AND TIME RESPONSE OF AUTOMATIC-CONTROL SYSTEMS

The transfer functions and time response of a system are characteristics for determining the system-output function, when the input function is known.

If the input and output signals are functions of time t, the system is described by its time response. If the signals are functions of the complex frequency s, the system is specified by its transfer function. We shall now consider the properties of these functions and the technique for their determination.

We remind the reader that the Laplace transform of a function of time $u(t)$ is a function $U(s)$ defined as

$$U(s)=\int_0^\infty u(t)e^{-st}\,dt. \tag{4.3}$$

In the Laplace transform the time t is replaced by the complex frequency

$$s=\sigma+j\omega, \tag{4.4}$$

where σ = the real part of the complex frequency;
ω = the imaginary part of this frequency.

The inverse transformation is obtained by solving (4.3) for $u(t)$:

$$u(t)=\frac{1}{2\pi j}\int_{-j\infty}^{j\infty} U(s)e^{ts}ds. \tag{4.5}$$

The Laplace transform applies the highly developed tools of the theory of functions of a complex variable to the analysis of automatic-control systems. This transformation enables differentiation to be replaced with multiplication by the complex frequency.

$$\frac{du}{dt}\to sU(s). \tag{4.6}$$

Differential equations are reduced to algebraic equations, greatly simplifying the solution of various problems.

Example. Find the equation for the current in the circuit of Figure 4.1 when the switch is closed (R and L are constants).

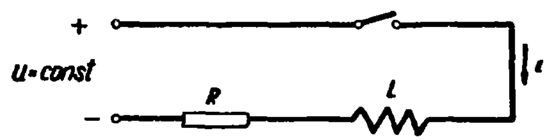

FIGURE 4.1. R-L circuit.

This circuit is described by the following equation:

$$u = Ri + L\frac{di}{dt}.$$

To find the current, we have to solve this differential equation. If, however, the equation is rewritten with the Laplace transforms for the current i and voltage u [$I(s)$ and $U(s)$ respectively], we obtain

$$U(s) = RI(s) + sLI(s).$$

This is a simple algebraic equation. The current (or rather its Laplace transform) is determined as the quotient

$$I(s) = \frac{U(s)}{R+sL}.$$

Since the voltage u is constant, its transform is equal to

$$U(s) = \frac{1}{s}u.$$

Hence

$$I(s) = \frac{1}{s(R+sL)}u = \frac{1}{s\left(s+\frac{1}{T}\right)}\frac{u}{L},$$

where $T = \frac{L}{R}$ is the time constant of the circuit

From tables of Laplace transforms we find the inverse function

$$i = T\left(1 - e^{-\frac{t}{T}}\right)\frac{u}{L} = \left(1 - e^{-\frac{t}{T}}\right)\frac{u}{R}.$$

Statistical theory covers the entire range of the function $u(t)$ from $t = -\infty$ to $t = \infty$. We therefore introduce the bilateral Laplace transform which, in distinction from the ordinary, unilateral transform (4.3), is defined as

$$U(s) = \int_{-\infty}^{\infty} u(t)e^{-st}\,dt. \tag{4.7}$$

The right-hand side of this equation can be written as a sum of two components:

$$U(s) = \int_{-\infty}^{0} u(t)e^{-st}\,dt + \int_{0}^{\infty} u(t)e^{-st}\,dt.$$

Substituting a variable $\tau=-t$ in the first integral we obtain

$$U(s)=-\int_{\infty}^{0}u(-\tau)e^{s\tau}\,d\tau+\int_{0}^{\infty}u(t)e^{-st}dt=$$
$$=\int_{0}^{\infty}u(-\tau)e^{s\tau}\,d\tau+\int_{0}^{\infty}u(t)e^{-st}dt.$$

We further introduce a new complex variable $s'=-s$ for the first integral. Thus,

$$U(s)=\int_{0}^{\infty}u(-\tau)e^{-s'\tau}d\tau+\int_{0}^{\infty}u(t)e^{-st}\,dt. \tag{4.8}$$

The bilateral Laplace transform of a function $u(t), \infty>t>-\infty$ is thus separated into two parts.

In the right integral of equation (4.8), we find the ordinary unilateral Laplace transform (4.3) of the function $u(t)$ defined for $t>0$.

Then, *in the left integral of* equation (4.8), we find the unilateral Laplace transform of the mirror image of the function $u(t)$ (Figure 4.2) and reverse the sign of the complex frequency in the transform obtained.

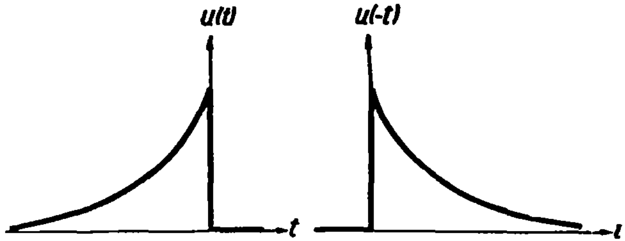

FIGURE 4.2. Mirror image (right) of the inverse function (left).

The sum of these two unilateral transforms gives the bilateral Laplace transform of the function $u(t)$

Example. *Find the bilateral Laplace transform of the function*

$$u(t)=\begin{cases} e^{at}; (t<0); \\ e^{-bt}; (t>0). \end{cases}$$

We first determine the unilateral Laplace transform of the function e^{-bt} defined for $t>0$. Using the Laplace-transform tables, we find

$$U_1(s)=\frac{1}{s+b}.$$

The second component is determined as follows:

a) the sign of t in $u(t)$ is reversed (we obtain the mirror image of the function $u(t)$ for $t<0$):

$$e^{at}\rightarrow e^{-at};$$

b) the Laplace transform of the mirror image of the function is according to (4.3):

$$\varepsilon^{-at} \to \frac{1}{s+a};$$

c) we change the sign of the complex frequency and obtain

$$U_2(s) = \frac{1}{-s+a} = -\frac{1}{s-a}.$$

The bilateral Laplace transform is thus equal to

$$U(s) = U_1(s) + U_2(s) = \frac{1}{s+b} - \frac{1}{s-a}.$$

In the complex plane (s) the automatic-control system is described by a transfer function which is defined as the ratio of the Laplace transform of the system's output function to the Laplace transform of the input function, with zero initial conditions:

$$G(s) = \frac{U_2(s)}{U_1(s)}. \tag{4.9}$$

Note that this function is independent of the shape of the input signal and is therefore suitable for statistical calculations. The transfer function varies with complex frequency $(s=\sigma+j\omega)$ and is plotted as a plane curve with coordinates σ (the real axis) and $j\omega$ (the imaginary axis).

If we set in (4.9) $s=j\omega$, we obtain the phase-amplitude response of the system:

$$G(j\omega) = \frac{U_2(j\omega)}{U_1(j\omega)}. \tag{4.10}$$

The substitution of the imaginary frequency $j\omega$ for the complex frequency s (is equivalent (see equation (4.4)) to setting the real frequency component σ equal to zero. Hence, the phase-amplitude response is a particular case of the transfer function tracing its variation along the imaginary frequency axis $(\sigma=0)$. The term "phase amplitude" is derived from the fact that when a sine signal is fed into the system this response function gives the phase shift and the amplitude of the output voltage, with respect to the input sine signal.

If the phase-amplitude response is known, then substituting s for $j\omega$ we obtain the transfer function.

The phase-amplitude response can be written in a somewhat modified form

$$G(j\omega) = B_1(\omega) + jB_2(\omega) = A(\omega)e^{j\upsilon(\omega)}, \tag{4.11}$$

where $A(\omega) = \sqrt{B_1^2(\omega) + B_2^2(\omega)}$;

$$\mathrm{tg}(\upsilon) = \frac{B_2(\omega)}{B_1(\omega)};$$

$B_1(\omega)$ and $B_2(\omega)$ = the real and the imaginary components of the phase-amplitude response.

The parameter $A(\omega)$ is often called the amplitude-frequency response and $\upsilon(\omega)$ the phase-frequency response.

Let us now consider an example in the determination of the transfer function and the frequency response.

Find the transfer function for a d-c generator (Figure 4. 3).

For the winding of this generator we have

$$u_1 = Ri_1 + L\frac{di_1}{dt}.$$

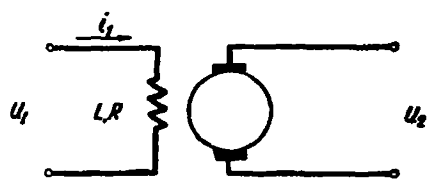

FIGURE 4. 3. Circuit of a d-c generator.

The Laplace tranform of this equation is

$$U_1(s) = (R + sL)I_1(s).$$

Neglecting saturation of the magnetic circuit of the generator, we write

$$u_2 = ki_1 \text{ or } U_2(s) = kI_1(s),$$

where k is a constant.

Then,

$$U_1(s) = \frac{1}{k}(R + sL)U_2(s), \tag{4.12}$$

whence the transfer function

$$G(s) = \frac{U_2(s)}{U_1(s)} = \frac{k}{R + sL}. \tag{4.13}$$

Setting $s = j\omega$ we obtain the open-loop frequency response function

$$G(j\omega) = \frac{k}{R + j\omega L}.$$

Using (4. 11), we write this expression in the form

$$G(j\omega) = \frac{k}{R}\left[\frac{1}{1+\omega^2T^2} - j\frac{\omega T}{1+\omega^2T^2}\right].$$

In Figure 4.4 we see that this function is a semicircle with its center at $B_1(\omega)=\frac{1}{2}\frac{k}{R}$ and $jB_2(\omega)=0$.

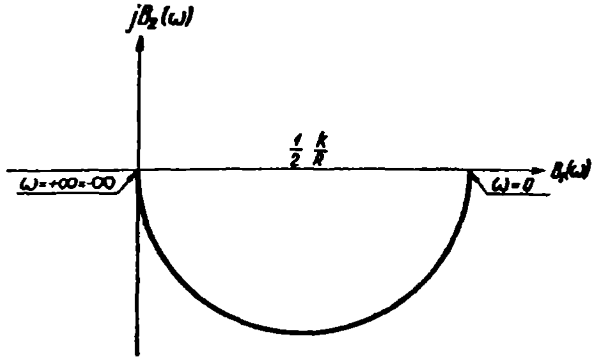

FIGURE 4.4. Complex plane plot of frequency response.

If the automatic-control system is closed (Figure 4.5), we may write

$$U_{\text{in}}(s)=U_1(s)-k_f U_2(s),$$

where k_f is the feedback transfer coefficient.

It is easy to see that the signal u_{in} in Figure 4.4 is analogous in its effect to u_1 in Figure 4.3. Therefore, from (4.12),

$$U_{\text{in}}(s)=\frac{1}{k}(R+sL)\,U_2(s).$$

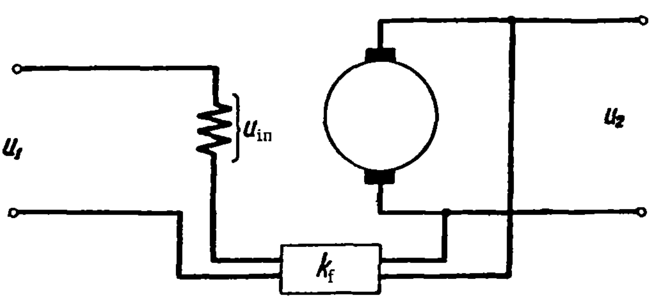

FIGURE 4.5. Closed-loop automatic-control system.

Solving these equations simultaneously, we obtain

$$U_1(s)=\left(k_f+\frac{R+sL}{k}\right)U_2(s),$$

whence we readily determine the transfer function of a closed-loop system:

$$G'(s)=\frac{U_2(s)}{U_1(s)}=\frac{k}{kk_f+R+sL}.$$

In general, the transfer function of a closed-loop system is obtained by inserting into this equation (see (4.13))

$$G_f\ (s)=k_f \text{ and } G(s)=\frac{k}{R+sL};$$

this yields

$$G'(s)=\frac{G(s)}{1+G(s)G_f\ (s)}. \qquad (4.14)$$

where $G_f\ (s)$ is the feedback transfer function.

All automatic-control systems, except for programed systems, are closed-loop and have one or more feedback loops.

In most cases automatic-control systems are studied in the complex plane. In some cases, however, it is convenient to make the analysis in the plane of the real variable t. We shall now therefore consider the relationships describing the system along the time axis.

Applying (4.5), we find the inverse Laplace transform for the transfer function:

$$g(t)=\frac{1}{2\pi j}\int_{-j\infty}^{j\infty}G(s)e^{ts}ds. \qquad (4.15)$$

The functions $g(t)$ and $G(s)$ are also related by the equation (4.3)

$$G(s)=\int_0^{\infty}g(t)e^{-st}dt. \qquad (4.16)$$

The parameter $g(t)$ is generally called the impulse time response of the system. It follows from equations (4.15) and (4.16) that the impulse time response and the impulse transfer function are related by Laplace transformation

$$g(t)\rightleftarrows G(s)$$

and characterize the system in the time and the frequency domains, respectively.

The output voltage of the system can be determined from the Laplace transform (4.5):

$$u_2=\frac{1}{2\pi j}\int_{-j\infty}^{j\infty}U_2(s)e^{ts}ds.$$

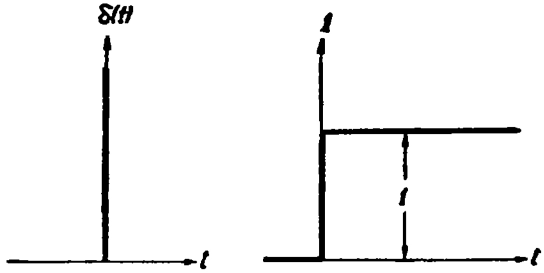

FIGURE 4. 6. Impulse (left) and unit-step (right) functions.

Since from (4. 9)

$$U_2(s)=G(s)\,U_1(s),$$

we have

$$u_2=\frac{1}{2\pi j}\int_{-j\infty}^{j\infty} G(s)\,U_1(s)\,e^{ts}ds.$$

In the particular case $U_1(s)=1$ we have

$$u_2=\frac{1}{2\pi j}\int_{-j\infty}^{j\infty} G(s)\,e^{ts}\,ds. \qquad (4.17)$$

Comparing equations (4. 15) and (4. 17) we come to the following conclusion: the impulse time response of the system $g(t)$ is equal to the output signal u_2 when the input of the system is a signal whose Laplace transform is $1\frac{1}{2}$.

Without proof we observe that the function whose Laplace transform is 1 is called an impulse or delta-function.

The impulse time response $g(t)$ of an automatic-control system is equal to the output signal when the input of the system is the impulse function $\delta(t)$.

The impulse function has the following property: at $t=0$ it is infinite, and at all other time points it is zero:

$$\delta(t)=\begin{cases}\infty & (t=0),\\ 0 & (t\neq 0).\end{cases} \qquad (4.18)$$

Moreover, the area described by the impulse function is

$$\int_{-\infty}^{\infty}\delta(t)\,dt=1.$$

The theory of automatic control often makes use of the unit-step function. This function (Figure 4.6) is defined as 1 for $t>0$ and 0 for $t<0$:

$$\mathbf{1}=\begin{cases}1 & (t>0),\\ 0 & (t<0).\end{cases} \qquad (4.19)$$

The relationship between the unit-step and the impulse function becomes evident when we differentiate the unit-step function:

$$\delta(t)=\frac{d(1)}{dt}. \tag{4.20}$$

We have shown previously that the impulse time response $g(t)$ is equal to the system-output signal when the input of the system is the impulse function. Similarly, a unit-step time response (or, more briefly, a unit-step response) $h(t)$ is introduced which is equal to the output signal of a system whose input is the unit-step function. It follows from (4.20) that these two time responses are related by the equation

$$g(t)=\frac{dh(t)}{dt}.$$

When the impulse function is applied to a system, its output is $g(t)$. What will be the output of a system whose input is an arbitrary signal curve?

We first find the bilateral Laplace transform for an arbitrary input function $u_1(t)$. It follows from (4.7) that

$$U_1(s)=\int_{-\infty}^{\infty} u_1(\sigma)\,e^{-s\sigma}\,d\sigma, \tag{4.21a}$$

where σ is introduced as a real time variable since later on several time variables will be required.

It moreover follows from (4.16) that the bilateral Laplace transform for the impulse time response is

$$G(s)=\int_{-\infty}^{\infty} g(\tau)e^{-s\tau}\,d\tau, \tag{4.21b}$$

where τ is another real time variable.

Inserting $U_1(s)$ and $G(s)$ into (4.9) we obtain the Laplace transform of the output function:

$$U_2(s)=G(s)\,U_1(s)=\int_{-\infty}^{\infty} g(\tau)e^{-s\tau}d\tau\int_{-\infty}^{\infty} u_1(\sigma)e^{-s\sigma}d\sigma=$$

$$=\int_{-\infty}^{\infty} g(\tau)d\tau\int_{-\infty}^{\infty} u_1(\sigma)e^{-s(\sigma+\tau)}d\sigma.$$

In the right-hand integral we introduce a new time variable

$$t=\tau+\sigma;$$

then

$$dt=d\sigma.$$

Thus,

$$U_2(s)=\int_{-\infty}^{\infty} g(\tau)d\tau\int_{-\infty}^{\infty} u_1(t-\tau)e^{-st}dt.$$

Changing the order of integration, we write

$$U_2(s)=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}u_1(t-\tau)g(\tau)d\tau\right]e^{-st}dt.$$

Comparing this equation with (4.7), we obtain

$$u_2(t)=\int_{-\infty}^{\infty}u_1(t-\tau)g(\tau)d\tau. \tag{4.22}$$

This integral can be interpreted as follows. The continuous input function $u_1(t)$ is resolved into an infinite number of impulses, and the integral sums the output signals produced, when these impulses are applied to the input.

Equation 4.22 is correspondingly called the convolution integral.

We have previously shown in (4.9) that in the complex-frequency domain the output signal is determined as the product of the input function and the transfer function

$$U_2(s)=G(s)U_1(s).$$

In the time domain this signal is determined by the convolution integral. We therefore say that multiplication in the frequency domain is equivalent to convolution in the time domain.

It is noteworthy that both τ and t in the convolution integral denote time. However, t is the time specifying the variation of the function $u_1(t)$, whereas τ is the time characterizing the shift of each of the infinite number of component impulses of this function relative to the ordinate axis. In computing the convolution integral, τ is considered as the variable, and the time t is constant, since integration is performed by summing the impulses over τ.

2. CORRELATION FUNCTIONS

We have already indicated that, in most cases, the object of statistical analysis is to minimize the mean square error (4.1).

We insert into (4.1) the output function (4.22). Hence

$$e^2=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\left[u_d(t)-\int_{-\infty}^{\infty}u_1(t-\tau)g(\tau)d\tau\right]^2dt.$$

Squaring the expression in brackets, we obtain

$$e^2=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\left[u_d^2(t)-2u_d(t)\int_{-\infty}^{\infty}u_1(t-\tau)g(\tau)d\tau+\right.$$

$$\left.+\int_{-\infty}^{\infty}u_1(t-\tau)g(\tau)d\tau\int_{-\infty}^{\infty}u_1(t-\sigma)g(\sigma)d\sigma\right]dt,$$

where σ is a new time variable to be distinguished from τ.

The mean square error is thus represented by three terms:

$$\overline{e^2} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} u_d^2(t)\,dt -$$

$$-2 \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} u_d(t)\,dt \int_{-\infty}^{\infty} u_1(t-\tau) g(\tau)\,d\tau +$$

$$+ \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \left[\int_{-\infty}^{\infty} u_1(t-\tau) g(\tau)\,d\tau \int_{-\infty}^{\infty} u_1(t-\sigma) g(\sigma)\,d\sigma \right] dt.$$

In the second and the third terms we change the order of integration and transition to the limit. Then

$$\overline{e^2} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} u_d^2(t)\,dt -$$

$$-2 \int_{-\infty}^{\infty} g(\tau)\,d\tau \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} u_d(t) u_1(t-\tau)\,dt +$$

$$+ \int_{-\infty}^{\infty} g(\tau)\,d\tau \int_{-\infty}^{\infty} g(\sigma)\,d\sigma \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} u_1(t-\tau) u_1(t-\sigma)\,dt. \qquad (4.23)$$

We introduce the following concepts:

a) autocorrelation function of the input signal:

$$\varphi_{11}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} u_1(t) u_1(t \pm \tau)\,dt, \qquad (4.24)$$

b) autocorrelation function of the output (the actual and the desired) signals:

$$\left.\begin{aligned} \varphi_{22}(\tau) &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} u_2(t) u_2(t \pm \tau)\,dt, \\ \varphi_{dd}(\tau) &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} u_d(t) u_d(t \pm \tau)\,dt, \end{aligned}\right\} \qquad (4.25)$$

c) cross-correlation function of the input and the output signals:

$$\left.\begin{aligned} \varphi_{12}(\tau) &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} u_1(t) u_2(t\pm\tau)\,dt, \\ \varphi_{1\mathrm{d}}(\tau) &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} u_1(t) u_{\mathrm{d}}(t\pm\tau)\,dt. \end{aligned}\right\} \tag{4.26}$$

The sign of τ in the integral is of no consequence and can be chosen arbitrarily (plus or minus).

It is easy to see that the correlation function is the average product over a period of time t of two functions, shifted one relative to the other by $\pm\tau$ seconds.

Inserting the correlation functions into (4.23) we obtain

$$e^2 = \varphi_{\mathrm{dd}}(0) - 2\int_{-\infty}^{\infty} \varphi_{1\mathrm{d}}(\tau) g(\tau)\,d\tau + $$

$$+ \int_{-\infty}^{\infty} g(\tau)\,d\tau \int_{-\infty}^{\infty} \varphi_{11}(\tau-\sigma) g(\sigma)\,d\sigma, \tag{4.27}$$

where $\varphi_{\mathrm{dd}}(0)$ is the autocorrelation function at $\tau=0$. Thus, the mean square error depends on the impulse time response and the correlation functions.

Note that when the impulse time response is known, the correlation functions can be expressed one in terms of the other. Let us consider these relationships. If an arbitrary signal is applied to the input, the output signal from equation (4.22) is equal to

$$u_2(t) = \int_{-\infty}^{\infty} u_1(t-\sigma) g(\sigma)\,d\sigma,$$

where σ and t are time variables.

Multiplying the two sides of this equation by $u_1(t-\tau)$, we obtain

$$u_1(t-\tau) u_2(t) = u_1(t-\tau) \int_{-\infty}^{\infty} u_1(t-\sigma) g(\sigma)\,d\sigma.$$

We now take the mean of the two sides of the equation, i.e., find the limit of these integrals:

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} u_1(t-\tau) u_2(t)\,dt =$$

$$= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \left[u_1(t-\tau) \int_{-\infty}^{\infty} u_1(t-\sigma) g(\sigma)\,d\sigma \right] dt.$$

The left-hand side of this equation, according to (4.26), is the cross-

correlation function. Therefore, seeing that $\varphi_{21}(\tau)=\varphi_{12}(\tau)$, we can write the above equation in the following form:

$$\varphi_{12}(\tau)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\left[u_1(t-\tau)\int_{-\infty}^{\infty}u_1(t-\sigma)g(\sigma)d\sigma\right]dt.$$

Changing the order of integration, we obtain

$$\varphi_{12}(\tau)=\int_{-\infty}^{\infty}g(\sigma)\left[\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}u_1(t-\tau)u_1(t-\sigma)dt\right]d\sigma.$$

The expression in brackets, according to (4.24), is an autocorrelation function of the argument $\tau-\sigma$. Therefore

$$\varphi_{12}(\tau)=\int_{-\infty}^{\infty}\varphi_{11}(\tau-\sigma)g(\sigma)d\sigma.$$

Since in this equation τ and σ are time variables, τ can be replaced by the more common t:

$$\varphi_{12}(\tau)=\int_{-\infty}^{\infty}\varphi_{11}(\tau-t)g(t)dt. \tag{4.28}$$

Thus, given the impulse time response of a system, we can obtain the cross-correlation function from the autocorrelation function of the input signal.

We further determine the relationship between the autocorrelation functions of the input and output signals.

The autocorrelation function of the output signal is from equation (4.25)

$$\varphi_{22}(\tau)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}u_2(t)u_2(t+\tau)dt.$$

From (4.22) we write

$$u_2(t)=\int_{-\infty}^{\infty}u_1(t-\sigma)g(\sigma)d\sigma;$$

$$u_2(t+\tau)=\int_{-\infty}^{\infty}u_1(t+\tau-\eta)g(\eta)d\eta,$$

where σ and η are time variables.

Inserting $u_2(t)$ and $u_2(t+\tau)$ into the preceding equation, we obtain

$$\varphi_{22}(\tau)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\left[\int_{-\infty}^{\infty}u_1(t-\sigma)g(\sigma)d\sigma\right]\left[\int_{-\infty}^{\infty}u_1(t+\tau-\eta)g(\eta)d\eta\right]dt.$$

Changing the order of integration, we write

$$\varphi_{22}(\tau)=\int_{-\infty}^{\infty} g(\sigma)d\sigma\int_{-\infty}^{\infty} g(\eta)d\eta\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} u_1(t+\tau-\eta)u_1(t-\sigma)dt.$$

Since from (4.24)

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} u_1(t+\tau-\eta)u_1(t-\sigma)dt=\varphi_{11}(\tau+\sigma-\eta),$$

we have

$$\varphi_{22}(\tau)=\int_{-\infty}^{\infty} g(\sigma)\left[\int_{-\infty}^{\infty}\varphi_{11}(\tau+\sigma-\eta)g(\eta)d\eta\right]d\sigma. \tag{4.29}$$

This expression establishes the relationship between the input and the output autocorrelation functions.

Let us consider some properties of correlation functions.

a) An autocorrelation function at $\tau=0$ characterizes the average power of the function. Indeed, let $u(t)$ be a voltage applied to a $1\,\Omega$ resistor. The autocorrelation functions equations (4.24) and (4.25) then give the input and the output power.

b) $|\varphi_{11}(\tau)|\leqslant\varphi_{11}(0)$. Let us prove this inequality. The product of two functions are expressed as follows:

$$\pm u_1(t)u_1(t+\tau)=\frac{1}{2}\left\{[u_1(t)\pm u_1(t+\tau)]^2-u_1^2(t)-u_1^2(t+\tau)\right\}.$$

Taking the mean, we obtain

$$\pm\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} u_1(t)u_1(t+\tau)dt=\frac{1}{2}\left\{\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}[u_1(t)\pm u_1(t+\tau)]^2dt-\right.$$

$$\left.-\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} u_1^2(t)dt-\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} u_1^2(t+\tau)dt\right\}.$$

Using the notations of (4.24), we write

$$\pm\varphi_{11}(\tau)=\frac{1}{2}\left\{\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}[u_1(t)\pm u_1(t+\tau)]^2dt-\varphi_{11}(0)-\varphi_{11}(0)\right\}$$

or

$$\mp\varphi_{11}(\tau)=\varphi_{11}(0)-\frac{1}{2}\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}[u_1(t)\pm u_1(t+\tau)]^2dt.$$

Since the second term in the right-hand side is always positive (the integral is squared), the absolute value of $\varphi_{11}(\tau)$ is always smaller than $\varphi_{11}(0)$.

c) If $u(t)$ contains a periodic or a constant component, the autocorrelation function also contains a periodic or a constant component. This follows from the very definition of correlation functions.

d) If $u_1(t)$ *is a random function, the cross-correlation function* $\varphi_{12}(\tau)\to 0$ when $\tau\to\infty$, since as τ increases the functions $u_1(t)$ and $u_2(t)$ become independent.

e) To a given autocorrelation function corresponds an infinite number of time functions, but to a given time function corresponds a single autocorrelation function. This is so, because the autocorrelation function is the mean of a function of time.

Let us now consider a technique for determining correlation functions using computing facilities. A block diagram of a correlation computor is shown in Figure 4.7.

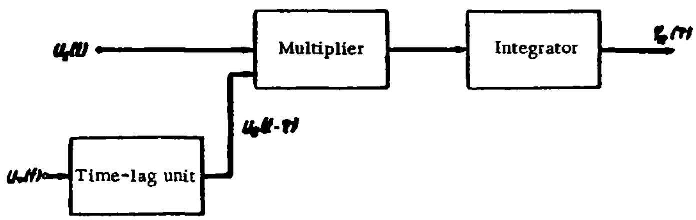

FIGURE 4.7. Correlation computor.

This computor uses analog devices and has two inputs: one for the input signal $u_1(t)$, and one for the output signal $u_2(t)$. Signal $u_2(t)$ is passed through a time-lag unit which produces a phase shift of τ seconds. The simplest time-lag unit is a magnetic tape with two read and write heads. If the speed of the tape is $V\frac{\text{cm}}{\text{sec}}$, the distance between the heads is

$$l=V\tau[\text{cm}].$$

Multiplication of the functions $u_1(t)$ and $u_2(t-\tau)$ and integration of the product gives, according to (4.26), the cross-correlation function. If the same signal is applied to the two inputs, the output is the autocorrelation function of this signal.

This facility is known to be inadequate when the integration limit $T\to\infty$. Therefore, in accordance with the law of large numbers, integration is made over a sufficiently large time period. The multiplier used in this facility should transmit high frequencies. Otherwise it will introduce a large error in the determination of the correlation function.

A circuit analogous to that in Figure 4.7 can be designed using digital devices, which perform numerical integration of the product of functions.

When the mathematical description of the functions $u_1(t)$ and $u_2(t)$ is given, the corresponding correlation functions can also be determined from equations (4.24), (4.25), and (4.26).

E x a m p l e. Find the autocorrelation function for the sine function

$$u_1(t) = u_m \sin(\omega t + \varphi).$$

According to (4.24), we write

$$\varphi_{11}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} u_m \sin(\omega t + \varphi)\, u_m \sin(\omega t + \varphi + \tau)\, dt.$$

Since $u_1(t)$ is a periodic function, it suffices to integrate over one period only:

$$\varphi_{11}(\tau) = \frac{1}{T} \int_{0}^{T} u_m \sin(\omega t + \varphi)\, u_m \sin(\omega t + \varphi + \tau)\, dt.$$

Since u_m = const and $T = \frac{2\pi}{\omega}$, we find

$$\varphi_{11}(\tau) = \frac{\omega}{2\pi} u_m^2 \int_{0}^{\frac{2\pi}{\omega}} \sin(\omega t + \varphi)\, \sin(\omega t + \varphi + \tau)\, dt.$$

Let

$$\vartheta = \omega t + \varphi.$$

Then $d\vartheta = \omega dt$ and

$$\varphi_{11}(\tau) = \frac{u_m^2}{2\pi} \int_{0}^{2\pi} \sin\vartheta\, \sin(\vartheta + \tau)\, d\vartheta,$$

whence, integrating, we obtain

$$\varphi_{11}(\tau) = \frac{1}{2} u_m^2 \cos\tau. \tag{4.30}$$

The autocorrelation function of a sine function is thus a cosine function which is independent of the origin of the time scale (the angle φ).

A method recently developed uses noise (generally considered an interference) for determining the impulse time response $g(t)$ of a system. This method is particularly valuable, as the response of the system can be determined without switching it off, and during its normal operation. Let us analyze this method.

W h i t e n o i s e — a perfectly random function (voltage) which is characterized by the lack of any correlation whatsoever between its consecutive values— is fed into the system. The autocorrelation function of white noise $w_1(t)$ is zero at all τ, with the exception of $\tau = 0$, where the random function

is multiplied by itself (see (4.24)). The autocorrelation function of white noise is therefore the impulse (or delta) function.

The relationship between the input and the output functions is determined by (4.22):

$$u_2(t)=\int_{-\infty}^{\infty} u_1(t-\tau)g(\tau)d\tau. \tag{4.31a}$$

If now, to facilitate comparison, the variables τ and t are interchanged in (4.28), we obtain

$$\varphi_{12}(t)=\int_{-\infty}^{\infty} \varphi_{11}(t-\tau)g(\tau)d\tau. \tag{4.31b}$$

In the previous section we showed that if the impulse function

$$u_1(t)=\delta(t),$$

is fed to the input of a system, (4.31a) is considerably simplified, and the output signal becomes equal to the impulse time response:

$$u_2(t)=g(t).$$

Since (4.31a) has the same form as (4.31b), we come to the following conclusion. If a signal (white noise) whose autocorrelation function is the impulse function

$$\varphi_{11}(t)=\delta(t),$$

is fed to the input, the cross-correlation function of the input and the output is equal to the impulse time response of the system:

$$\varphi_{12}(t)=g(t).$$

We thus have the possibility of designing a circuit (Figure 4.8) for determining the impulse time response of the system.

Let us analyze the operation of the circuit shown in Figure 4.8.

White noise $w_1(t)$ from a special generator is delivered simultaneously to the system input, and to a time-lag unit. The time-lag unit produces a phase shift of τ sec in $w_1(t)$. When the noise has passed through the system, signal $w_2(t)$ is produced at the output. The functions $w_1(t-\tau)$ and $w_2(t)$ are fed into a correlation computor which produces an output signal equal to the impulse time response $g(t)$ of the system.

If the input of the system consists not only of white noise, but also of the signal $u_1(t)$, the computor receives the sum of the signals

$$w_2(t)+u_2(t),$$

where $u_2(t)$ is the output produced by the system when the signal $u_1(t)$ has passed through. But the cross-correlation function of $u_1(t)$ and $u_2(t)$ is zero, since $u_1(t)$ is not fed into the computor, and its signal is therefore zero.

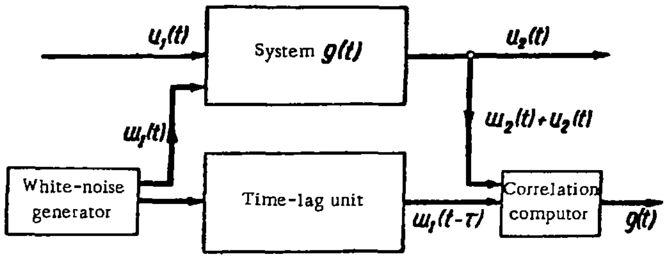

FIGURE 4. 8. Circuit for determining the impulse time response of a system.

The circuit in Figure 4. 8 thus enables us to determine the impulse time response of a system for any external input signal $u_1(t)$. It is therefore possible to measure $g(t)$ directly during operation, without switching off the system. Moreover, external noise introduced into the system or generated in it does not affect the measurements of $g(t)$, since it introduces no errors.

In minimizing the mean square error (4. 27) of a system, the correlation functions are used to describe the signals fed into the system and its behavior in time. Forming the bilateral Laplace transform (4. 7) of the correlation functions, we obtain the characteristic responses in the complex-frequency domain:

$$\left.\begin{aligned}\Phi_{11}(s)&=\int_{-\infty}^{\infty}\varphi_{11}(\tau)e^{-s\tau}d\tau;\\ \Phi_{12}(s)&=\int_{-\infty}^{\infty}\varphi_{12}(\tau)e^{-s\tau}d\tau;\\ \Phi_{22}(s)&=\int_{-\infty}^{\infty}\varphi_{12}(\tau)e^{-s\tau}d\tau.\end{aligned}\right\}\quad(4.32)$$

$\Phi_{11}(s)$ is not the only Laplace transform of the autocorrelation function; this function in its own right gives a very important frequency response. Let us consider this property of $\Phi_{11}(s)$.

Any continuous function $u(t)$ can be expanded in any finite interval (T) into Fourier series, representing a sum of individual harmonics (sine and cosine curves of different amplitudes and frequencies):

$$\begin{aligned}u(t)=\frac{a_0}{2}&+a_1\cos\omega t+a_2\cos 2\omega t+\ldots+a_n\cos n\omega t+\ldots+\\&+b_1\sin\omega t+b_2\sin 2\omega t+\ldots+b_n\sin n\omega t+\ldots,\end{aligned}\quad(4.33)$$

where

$$a_n=\frac{2}{T}\int_0^T u(t)\cos n\omega t\,dt;$$

$$b_n=\frac{2}{T}\int_0^T u(t)\sin n\omega t\,dt.$$

The complex Fourier series is written in a somewhat different form:

$$u(t)=\frac{1}{T}\sum_{n=-\infty}^{\infty}u_n(j\omega)\,e^{jn\omega t}, \tag{4.34}$$

where

$$u_n(j\omega)=\int_{-\frac{T}{2}}^{\frac{T}{2}}u(t)\,e^{-jn\omega t}dt. \tag{4.35}$$

It follows from (4.34) *that function* $u(t)$ can be represented as an *infinite* discrete series of harmonics whose amplitudes are determined by (4.35).

Let a voltage $u(t)$ be applied to a circuit with resistance R. Each harmonic component of this voltage causes to dissipate in a resistor R an amount of energy equal to

$$\frac{u_n^2(j\omega)}{R}.$$

Since the spectrum of the voltage function $u(t)$ is discrete, (4.34) we say that energy in signal $u(t)$ is concentrated at discrete (individual) frequencies.

It follows from (4.35) that the frequency as a function of the harmonic amplitudes (the amplitude spectrum) is represented by a discrete curve (Figure 4.9), and is defined only at points where n is an integer.

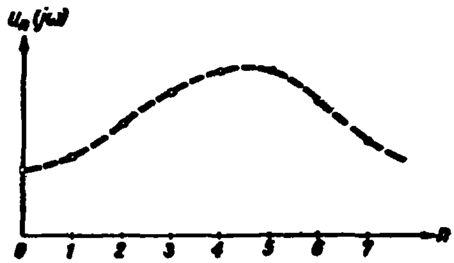

FIGURE 4. 9. Discrete amplitude spectrum.

Until now we have considered the expansion of the $u(t)$ curve in the time interval T. If, for a periodic curve, this interval is taken equal to the period of the function, the curve will repeat itself outside this interval. It is therefor exactly specified by its discrete spectrum.

Aperiodic functions, on the other hand, are approximated (with a certain error) by a discrete spectrum in an interval of T seconds. To reduce this error, the interval T must be increased. In the limit, when $T=\infty$, the

error vanishes, and the Fourier series of equation (4.34) reduces to the Fourier integral:

$$u(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} u(j\omega)\, e^{j\omega t}d\omega, \tag{4.36}$$

where

$$u(j\omega)=\int_{-\infty}^{\infty} u(t)\, e^{-j\omega t}dt. \tag{4.37}$$

The continuous function $u(j\omega)$ is the limit of the discrete sequence $u_n(j\omega)$ (equation (4.35) and Figure 4.9) as the difference between these values approaches zero. It follows from (4.37) that in this case the voltage $u(t)$ has a continuous spectrum of harmonics and the energy content of the function $u(t)$ is continuously distributed over the entire frequency range.

In the discrete spectrum, the function $u_n(j\omega)$ describes the frequency as a function of the amplitudes of the harmonics, whereas in the continuous spectrum the function $u(j\omega)$ describes the frequency dependence of the amplitude density of the harmonics (i.e., the ratio of the amplitudes to an infinitesimal frequency interval) (Figure 4.10). We have already encountered a similar phenomenon in our analysis of the differential probability-distribution function (see page 106). There, the transition from a discrete to a continuous function resulted in a switch from probability to probability density, defined as the ratio of probability to an infinitesimal signal interval.

Let us now find the direct relationship of $\Phi_{11}(s)$ (4.32) with a signal $u_1(t)$, whose autocorrelation function $\varphi_{11}(\tau)$ is known.

From (4.24)

$$\varphi_{11}(\tau)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} u_1(t)u_1(t+\tau)dt.$$

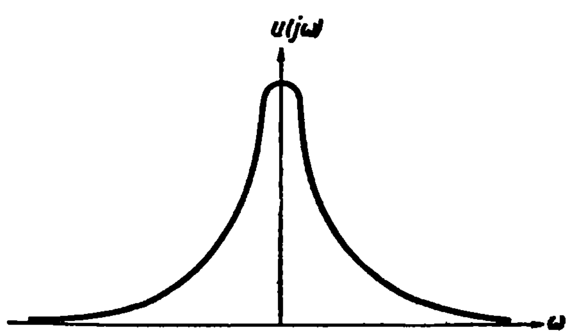

FIGURE 4.10. Amplitude density of a continuous spectrum.

Inserting $\varphi_{11}(\tau)$ into (4.32), we obtain

$$\Phi_{11}(s)=\int_{-\infty}^{\infty} e^{-s\tau}d\tau \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} u_1(t)u_1(t+\tau)dt.$$

Changing the order of integration, we write

$$\Phi_{11}(s)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} u_1(t)dt\int_{-\infty}^{\infty} u_1(t+\tau)e^{-s\tau}d\tau.$$

Proceeding from the law of large numbers, we substitute finite limits for the infinite limits of integration (T is sufficiently large). Then

$$\Phi_{11}(s)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} u_1(t)dt\int_{-T}^{T} u_1(t+\tau)e^{-s\tau}d\tau.$$

Changing the variable τ to $\sigma=t+\tau$, so that $d\sigma=d\tau$ and noting that as $T\to\infty, \pm T+t\approx\pm T$, we obtain

$$\Phi_{11}(s)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} u_1(t)dt\int_{-T}^{T} u_1(\sigma)e^{-s(\sigma-t)}\,d\sigma$$

or

$$\Phi_{11}(s)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} u_1e^{st}\,dt\int_{-T} u_1(\sigma)e^{-s\sigma}\,d\sigma.$$

The two integrals are conjugate. Therefore

$$\Phi_{11}(s)=\lim_{T\to\infty}\frac{1}{2T}\left|\int_{-T}^{T} u_1(t)e^{-st}\,dt\right|^2 \tag{4.38}$$

or

$$\Phi_{11}(j\omega)=\lim_{T\to\infty}\frac{1}{2T}\left|\int_{-T}^{T} u_1(t)e^{-j\omega t}dt\right|^2. \tag{4.39}$$

It follows from (4.37) that the right-hand side of (4.39) is proportional to the square of the amplitude density. The function $\Phi_{11}(j\omega)$ is therefore called **the spectral power density**. It is easy to see that $\Phi_{11}(j\omega)$ is a real function of the square of frequency (ω^2).

A periodic signal is thus characterized by a discrete amplitude spectrum, and an aperiodic signal by a continuous amplitude-density spectrum; a random signal, on the other hand, is characterized by spectral power density.

Spectral density can also characterize the distribution of power in the frequency spectrum of periodic and aperiodic signals.

The function $\Phi_{11}(j\omega)$ gives the distribution of power density over the entire harmonics spectrum. The power dissipated on a resistor $R=1\,\Omega$ by harmonics having frequencies from ω_1 to ω_2 is therefore determined by the integral

$$W_{\omega_1-\omega_2}=\frac{1}{2\pi}\int_{\omega_1}^{\omega_2}\Phi_{11}(j\omega)d\omega. \tag{4.40}$$

The total power content of the signal $u_1(t)$ (i.e., the power transmitted by all the frequencies) is equal to

$$W=\frac{1}{2\pi}\int_{-\infty}^{\infty}\Phi_{11}(j\omega)d\omega. \tag{4.41}$$

It should be remembered that $\Phi_{11}(j\omega)$ is a response specifying the spectral power density of the signal $u_1(t)$. It contains no amplitude frequency responses.

By analogy with $\Phi_{11}(j\omega)$, the function $\Phi_{12}(j\omega)$ (4.32) is called the c r o s s s p e c t r a l p o w e r d e n s i t y.

We have previously shown in equation (4.10) that in the complex-frequency domain the input and the output signals of a system are related by the expression

$$U_2(j\omega)=G(j\omega)U_1(j\omega).$$

An analogous relationship in the time domain was given by the convolution integral (4.22):

$$u_2(\tau)=\int_{-\infty}^{\infty}u_1(\tau-t)g(t)dt.$$

Correlation functions in the time domain have the same general form (see equation (4.28)):

$$\varphi_{12}(\tau)=\int_{-\infty}^{\infty}\varphi_{11}(\tau-t)g(t)dt.$$

Laplace transforms of the correlation functions are therefore also related by the dependence

$$\Phi_{12}(j\omega)=G(j\omega)\Phi_{11}(j\omega). \tag{4.42}$$

Let us now determine a relationship between the spectral densities of the input $[\Phi_{11}(j\omega)]$ and output $[\Phi_{22}(j\omega)]$ signals.

Inserting the autocorrelation function of the output signal from (4.29) into (4.32), we obtain

$$\Phi_{22}(j\omega)=\int_{-\infty}^{\infty} e^{-j\omega\tau}\left[\int_{-\infty}^{\infty} g(\sigma)d\sigma\int_{-\infty}^{\infty}\varphi_{11}(\tau+\sigma-\eta)g(\eta)d\eta\right]d\tau=$$

$$=\int_{-\infty}^{\infty} d\tau\int_{-\infty}^{\infty} g(\sigma)d\sigma\times$$

$$\times\int_{-\infty}^{\infty} e^{-j\omega(\tau+\sigma-\eta)}e^{j\omega\sigma}e^{-j\omega\eta}\,\varphi_{11}(\tau+\sigma-\eta)g(\eta)d\eta.$$

Changing the order of integration, we have

$$\Phi_{22}(j\omega)=\int_{-\infty}^{\infty} g(\eta)e^{-j\omega\eta}d\eta\int_{-\infty}^{\infty} g(\sigma)e^{j\omega\sigma}d\sigma\times$$

$$\times\int_{-\infty}^{\infty}\varphi_{11}(\tau+\sigma-\eta)e^{-j\omega(\tau+\sigma-\eta)}\,d\tau.$$

Substituting $j\omega$ for s in (4.16) and considering the entire time range $(\infty>t>-\infty)$, we write

$$G(j\omega)=\int_{-\infty}^{\infty} g(t)e^{-j\omega t}dt.$$

The preceding equation can therefore be rewritten as

$$\Phi_{22}(j\omega)=G(j\omega)G(-j\omega)\int_{-\infty}^{\infty}\varphi_{11}(\lambda)e^{-j\omega\lambda}\,d\lambda,$$

where $\lambda=\tau+\sigma-\eta;$
$d\lambda=d\tau.$

Since the last integral is in fact the spectral power density of the input signal (4.32), the equation relating the spectral densities of the input and the output signals can therefore be written as

$$\Phi_{22}(j\omega)=|G(j\omega)|^2\Phi_{11}(j\omega), \tag{4.43}$$

where $|G(j\omega)|^2=G(j\omega)G(-j\omega)$.

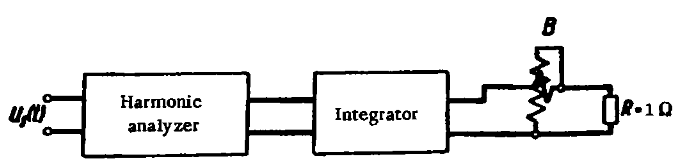

FIGURE 4.11. Circuit for the determination of the spectral power density.

The spectral power density of the input signal $\Phi_{11}(j\omega)$ can be directly determined by experiment. Since $\Phi_{11}(j\omega)$ is equal to the power of the signal $u_1(t)$ at frequencies between ω and $\omega+d\omega$, the density can be determined as follows.

The random function in question, $u_1(t)$, is fed into a harmonic analyzer (Figure 4.11). This device is a narrow-band filter admitting only a small spectrum of harmonics. Since the transmission band is nevertheless not infinitesimal, the output of the analyzer is fed into an integrator which averages out the harmonics spectrum. The output of the integrator is coupled to a resistance $R=1\,\Omega$. A wattmeter (B) measures the power dissipated in the resistor in this frequency band. The power measured gives a point on the $\Phi_{11}(j\omega)$ curve. Returning the analyzer, we determine in this way several points for plotting the curve $\Phi_{11}(j\omega)=f(\omega)$.

When the voltage $u_1(t)$ is given analytically, the spectral power density can be computed from equations (4.32) or (4.39).

As we have previously shown, the autocorrelation function of white noise is the impulse function

$$\varphi_{11}(\tau)=\delta(\tau).$$

Therefore, inserting $\varphi_{11}(\tau)$ into (4.32), we obtain

$$\Phi_{11}(j\omega)=\int_{-\infty}^{\infty}\delta(\tau)e^{-j\omega\tau}\,d\tau. \tag{4.44}$$

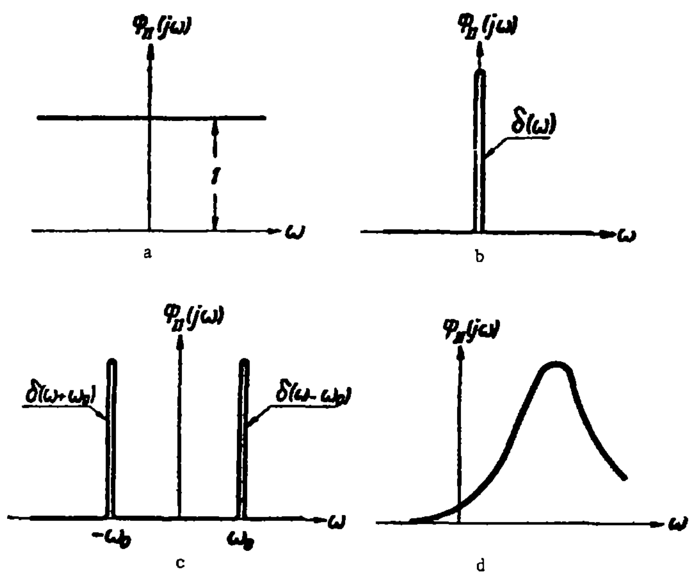

FIGURE 4.12. Spectral power density: a-white noise; b-constant signal; c-periodic signal; d-random signal.

Integrating, we find

$$\Phi_{11}(j\omega) = 1.$$

Different types of spectral densities for the most common functions are shown in Figure 4.12. We see from this figure that the spectral power density of white noise is a horizontal straight line with ordinate 1. The spectral power density of a constant signal is represented by the impulse function at the origin. Two impulse functions at the abscissas $\omega_0 = \pm \frac{2\pi}{T}$ characterize a periodic sinusoidal signal with frequency ω_0.

The spectral power density, $\Phi_{11}(j\omega)$, of a random signal is a smooth curve.

3. MINIMIZING THE MEAN SQUARE ERROR

The principal problem encountered in the design of automatic-control systems is the design of a system (how to choose $G(s)$ or $g(t)$) whose mean square error at the output will be minimized.

The mean square error can be determined from the impulse time response and the correlation functions (see (4.27)). It is, however, much simpler to determine this error from the transfer functions of the system in the complex-frequency domain. It should be kept in mind that the mean square error consists of two components: errors arising in the transmission of noise $w_1(t)$ through the systems and errors in the transmission of the meaningful signal $u_1(t)$. Since we are concerned with linear systems only, each of the component errors can be computed separately. We shall therefore consider two cases of system operation, with noise and meaningful signal fed separately into the system.

a) *Transmission of noise*

In this case (Figure 4.13) noise only is fed into the sytem $[w_1(t)]$. The output signal $w_2(t)$ is obviously meaningless and thus constitutes an error.

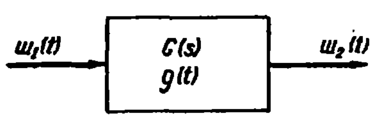

FIGURE 4.13. Transmission of noise.

The total power of the output signal, and hence the mean square error is, from equation (4.41), equal to

$$e_m^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_{ns_{22}}(j\omega)\, d\omega,$$

where $\Phi_{ns_{2}}$ is the spectral power density of the output noise.

Inserting the expression for spectral power density of the output signal from (4.43), we obtain

$$e^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} |G(j\omega)|^2 \, \Phi_{ns_{n}}(j\omega)\,d\omega, \tag{4.45}$$

where $\Phi_{ns_{n}}$ $(j\omega)$ is the spectral power density of the input noise.

b) Error in transmission of meaningful signal

This error arises from two sources. First, the system available may process the signal somewhat differently from the initial specifications and, second, the system may differ from the optimum if the mean square error is to be minimized.

It is convenient to use the circuit shown in Figure 4.14 for the analysis of the error arising in the transmission of a meaningful signal. This circuit gives the error as a difference of signals at the outputs of two systems:

$$e(t) = u_{\mathrm{d}}(t) - u_2(t). \tag{4.46}$$

One of these systems is the physically available real system $G(s)$, $g(t)$, and the other is the desired system $G_{\mathrm{d}}(s)$, $g_{\mathrm{d}}(t)$.

From (4.22), the output signal of the real system is

$$u_2(t) = \int_{-\infty}^{\infty} u_1(t-\tau)g(\tau)\,d\tau.$$

Similarly, the output signal of the desired system is

$$u_{\mathrm{d}}(t) = \int_{-\infty}^{\infty} u_1(t-\tau)g_{\mathrm{d}}(\tau)\,d\tau.$$

Inserting signals $u_2(t)$ and $u_{\mathrm{d}}(t)$ into (4.46), we obtain

$$e(t) = \int_{-\infty}^{\infty} u_1(t-\tau)[g_{\mathrm{d}}(\tau) - g(\tau)]\,d\tau.$$

The error $e(t)$ can thus be considered (Figure 4.15) as the output signal of a conditional system $[G_{\mathrm{d}}(s) - G(s); g_{\mathrm{d}}(t) - g(t)]$, whose input is the meaningful signal $u_1(t)$. The mean square error arising in the transmission of a meaningful signal can thus be determined as the integral power of the signal produced at the output of the conditional system. Therefore, similar to (4.45), we obtain

$$e_n^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} |G_{\mathrm{d}}(j\omega) - G(j\omega)|^2 \Phi_{s_{n}}(j\omega)\,d\omega, \tag{4.47}$$

where $G_d(j\omega)$ = the frequency response of the desired system;
$G(j\omega)$ = the frequency response of the real system;
$\Phi_{11}(j\omega)$ = the spectral power density of the meaningful input signal.

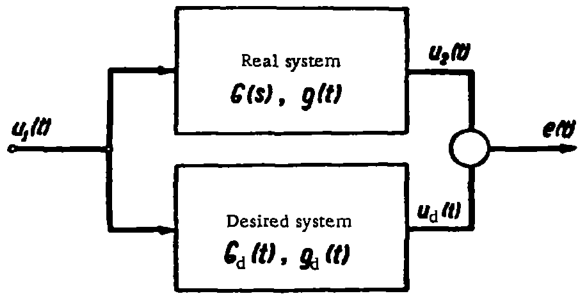

FIGURE 4.14. Circuit for determining the error in transmission of meaningful signal.

The total error measured at the output is made up of noise error plus the transmission error of the meaningful signal:

$$e^2 = e_{ns}^2 + e_n^2.$$

Inserting the expression for these errors as given by equations (4.45) and (4.47) we obtain

$$e^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[|G(j\omega)|^2\Phi_{ns_n}(j\omega) + |G_d(j\omega) - G(j\omega)|^2\Phi_{n_n}(j\omega)\right]d\omega. \qquad (4.48)$$

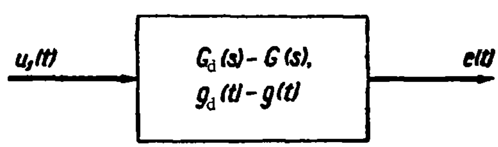

FIGURE 4.15. Transmission of a signal through a conditional system.

Example. Find the mean square error at the output of the circuit shown in Figure 4.16. The input of the system is white noise with spectral density $\Phi_{ns_n}(j\omega) = 2$. The circuit parameters: $R = 500\,\Omega$, $C = 0.01$ F.

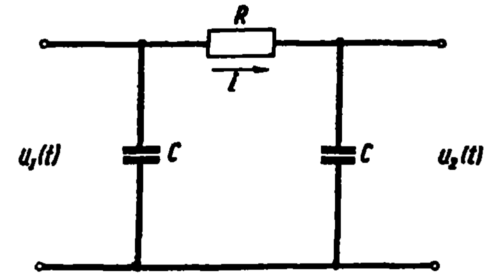

FIGURE 4.16. Schematic diagram of a system.

We first determine the transfer function of the system. For the circuit in Figure 4.16 we may write

$$U_1(s)=\left(R+\frac{1}{sC}\right)I(s);$$

$$U_2(s)=\frac{1}{sC}I(s).$$

Solving these equations simultaneously, we derive the transfer function

$$G(s)=\frac{U_2(s)}{U_1(s)}=\frac{1}{sCR+1}=\frac{1}{CR}\cdot\frac{1}{s+\frac{1}{CR}}.$$

Inserting the circuit parameters into this equation, we obtain

$$G(s)=\frac{1}{5}\cdot\frac{1}{s+0.2}$$

or

$$G(j\omega)=\frac{1}{5}\cdot\frac{1}{j\omega+0.2}.$$

The mean square error at the output is according to (4.45) equal to

$$e^2=\frac{1}{2\pi}\int_{-\infty}^{\infty}|G(j\omega)|^2\Phi_{ns_M}(j\omega)\,d\omega.$$

Inserting the parameters, we have

$$e^2=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\frac{1}{5}\cdot\frac{1}{j\omega+0.2}\right|^2 2d\omega=\frac{1}{25\pi}\int_{-\infty}^{\infty}\left|\frac{1}{j\omega+0.2}\right|^2 d\omega.$$

Seeing (4.43) that

$$|G(j\omega)|^2 = G(j\omega)\,G(-j\omega),$$

we find

$$e^2 = \frac{1}{25\pi}\int_{-\infty}^{\infty}\frac{1}{(j\omega+0.2)(-j\omega+0.2)}\,d\omega = \frac{1}{25\pi}\int_{-\infty}^{\infty}\frac{1}{\omega^2+0.04}\,d\omega.$$

Integrating, we obtain

$$e^2 = \frac{1}{25\pi}\cdot\frac{1}{\sqrt{0.04}}\left[\operatorname{arctg}\frac{\omega}{\sqrt{0.04}}\right]_{-\infty}^{\infty} = 0.2.$$

Using the same example, we shall now determine the spectral density of the output signal.

It follows from (4.43) that the spectral density of the output signal is equal to

$$\Phi_{ns_{22}}(j\omega) = |G(j\omega)|^2\,\Phi_{ns_{11}}(j\omega).$$

Inserting the parameters, we obtain

$$\Phi_{ns_{22}}(j\omega) = \frac{1}{25}\cdot\frac{1}{\omega^2+0.04}\,2 = \frac{0.08}{\omega^2+0.04}.$$

To optimize an automatic-control system, it suffices to determine the transfer function of a system whose output mean square error is minimum. Moreover, the constraints imposed by the condition of the system being physically stable must be taken into consideration. This condition requries that the system remain at rest (output signal be zero) as long as no signal is fed to the input. Mathematically, this condition is written as

$$g(t) = 0 \text{ when } t < 0.$$

Without proof, we state that for a system to be stable, its transfer function must be analytical in the right half of the complex-frequency plane.

In this case all the poles* of the transfer function fall in the left half-plane.

Let us first consider how the mean square error is minimized, without considering the constraints imposed by the condition of stability.

* A pole of a function is a point at which the function assumes an infinite value. For example, the point $s = -a$ is a pole of the function $U(s) = \frac{1}{s+a}$.

Take the expression for the phase-amplitude response in (4.11)

$$G_d(j\omega)=A_d(\omega)e^{j\upsilon_d(\omega)}=A_d(\omega)[\cos\upsilon_d(\omega)+j\sin\upsilon_d(\omega)], \tag{4.49}$$

where $A_d(\omega)$ and $\upsilon_d(\omega)$ are the modulus and the phase of the given phase-amplitude response (real functions of frequency). To minimize the mean square error at the output, we modify the phase-amplitude response and set it at

$$G_{\text{opt}}(j\omega)=A_{\text{opt}}(\omega)e^{j\upsilon_{\text{opt}}(\omega)}=A_{\text{opt}}(\omega)[\cos\upsilon_{\text{opt}}(\omega)+j\sin\upsilon_{\text{opt}}(\omega)], \tag{4.50}$$

where $A_{\text{opt}}(\omega)$ and $\upsilon_{\text{opt}}(\omega)$ are the modulus and the phase of the optimum phase-amplitude response. The mean square error at the system output is then, according to (4.48), equal to

$$e^2=\frac{1}{2\pi}\int_{-\infty}^{\infty}[|G_{\text{opt}}(j\omega)|^2\,\Phi_{\text{ns}_{11}}(j\omega)+|G_d(j\omega)-G_{\text{opt}}(j\omega)|^2\,\Phi_{n_{11}}(j\omega)]\,d\omega. \tag{4.51}$$

Inserting the phase-amplitude responses from equations (4.49) and (4.50) into this equation we obtain (the arguments ω and $j\omega$ have been omitted for brevity):

$$e^2=\frac{1}{2\pi}\int_{-\infty}^{\infty}\Big[A_{\text{opt}}^2e^{j\upsilon_{\text{opt}}}e^{-j\upsilon_{\text{opt}}}\Phi_{\text{ns}_{11}}+|(A_d\cos\upsilon_d-A_{\text{opt}}\cos\upsilon_{\text{opt}})+$$

$$+j(A_d\sin\upsilon_d-A_{\text{opt}}\sin\upsilon_{\text{opt}})|^2\,\Phi_{n_{11}}\Big]\,d\omega. \tag{4.52}$$

Now,

$$\begin{aligned}
&|(A_d\cos\upsilon_d-A_{\text{opt}}\cos\upsilon_{\text{opt}})+j(A_d\sin\upsilon_d-A_{\text{opt}}\sin\upsilon_d)|^2=\\
&=[(A_d\cos\upsilon_d-A_{\text{opt}}\cos\upsilon_{\text{opt}})+j(A_d\sin\upsilon_d-A_{\text{opt}}\sin\upsilon_d)]\times\\
&\times[(A_d\cos\upsilon_d-A_{\text{opt}}\cos\upsilon_{\text{opt}})-j(A_d\sin\upsilon_{\text{opt}}-A_{\text{opt}}\sin\upsilon_d)]=\\
&=(A_d\cos\upsilon_d-A_{\text{opt}}\cos\upsilon_{\text{opt}})^2+(A_d\sin\upsilon_{\text{opt}}-A_{\text{opt}}\sin\upsilon_d)^2=\\
&=A_d^2+A_{\text{opt}}^2-2A_{\text{opt}}A_d(\cos\upsilon_d\cos\upsilon_{\text{opt}}+\sin\upsilon_d\sin\upsilon_{\text{opt}})=\\
&=A_d^2+A_{\text{opt}}^2-2A_{\text{opt}}A_d\cos(\upsilon_{\text{opt}}-\upsilon_d).
\end{aligned}$$

Therefore, seeing that $e^{j\upsilon_{\text{opt}}}e^{-j\upsilon_{\text{opt}}}=1$, we rewrite (4.52) in the form

$$e^2=\frac{1}{2\pi}\int_{-\infty}^{\infty}\Big\{A_{\text{opt}}^2\,\Phi_{\text{ns}_{11}}+[A_d^2+A_{\text{opt}}^2-2A_{\text{opt}}A_d\cos(\upsilon_{\text{opt}}-\upsilon_d)\Phi_{n_{11}}]\Big\}\,d\omega.$$

The coefficients A_{opt}, A_d, $\Phi_{\text{ns}_{11}}$, and $\Phi_{n_{11}}$ are positive for all frequencies ω.

Therefore, to minimize the mean square error we must maximize the negative term $2A_{opt}A_d\cos(\upsilon_{opt}-\upsilon_d)$, i.e., make

$$\upsilon_{opt}=\upsilon_d.$$

Hence, to minimize the mean square error, the phase of the phase-amplitude response need not be changed, and may remain equal to the given phase υ_d.

Further, since $\upsilon_{opt}=\upsilon_d$ (4.53), $\cos(\upsilon_{opt}-\upsilon_d)=1$, we can rewrite the preceding equation in the following form:

$$e^2=\frac{1}{2\pi}\int_{-\infty}^{\infty}[A_{opt}^2\Phi_{ns_{11}}+(A_d^2+A_{opt}^2-2A_{opt}A_d)\Phi_{n_{11}}]\,d\omega$$

or

$$e^2=\frac{1}{2\pi}\int_{-\infty}^{\infty}[A_{opt}^2(\Phi_{n_{11}}+\Phi_{ns_{11}})-2A_{opt}A_d\Phi_{n_{11}}+A_d^2\Phi_{n_{11}}]\,d\omega.$$

Since all the terms in the integral are positive, the minimum mean square error is determined by the function

$$F(A_{opt})=A_{opt}^2(\Phi_{n_{11}}+\Phi_{ns_{11}})-2A_{opt}A_d\Phi_{n_{11}}+A_d^2\Phi_{n_{11}}.$$

The parameter A_{opt} is thus determined from the equation

$$\frac{dF(A_{opt})}{dA_{opt}}=0.$$

Differentiating $F(A_{opt})$ and setting the derivative equal to zero, we obtain

$$A_{opt}=\frac{\Phi_{n_{11}}}{\Phi_{n_{11}}+\Phi_{ns_{11}}}A_d. \tag{4.54}$$

Choosing the parameters A_{opt} and υ_{opt} from equations (4.53) and (4.54) we thus minimize the mean square error:

$$e^2=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\Phi_{n_{11}}\Phi_{ns_{11}}}{\Phi_{n_{11}}+\Phi_{ns_{11}}}A_d^2\,d\omega. \tag{4.55}$$

The optimum frequency response is determined by dividing (4.50) by (4.49)

$$\frac{G_{opt}(j\omega)}{G_d(j\omega)}=\frac{A_{opt}e^{j\upsilon_{opt}}}{A_de^{j\upsilon_d}}.$$

Inserting υ_{opt} from (4. 53) and A_{opt} from (4. 54) gives

$$G_{\text{opt}}(j\omega)=\frac{\Phi_{n_{\text{u}}}(j\omega)}{\Phi_{n_{\text{u}}}(j\omega)+\Phi_{\text{ns}_{\text{u}}}(j\omega)}G_d(j\omega). \tag{4. 56a}$$

If s is substituted for $j\omega$, the optimum transfer function is

$$G_{\text{opt}}(s)=\frac{\Phi_{n_{\text{u}}}(s)}{\Phi_{n_{\text{u}}}(s)+\Phi_{\text{ns}_{\text{u}}}(s)}G_d(s). \tag{4. 56b}$$

It follows from equations (4. 56a)and (4. 56b)that in the absence of input noise ($\Phi_{\text{ns}_{\text{u}}}=0$) the optimum transfer function (and frequency response) is equal to the given function. The optimization, thus in fact, reduces the effect of noise on the output function of the system. If the system is noiseless, there is no point in this optimization.

We may thus say that (4. 56b) determines the optimum transfer function $G_{\text{opt}}(s)$, of a system with noise, from the given transfer function $G_d(s)$, of the noiseless system.

Consider an example in the determination of an optimum transfer function.

Let the spectral densities of the meaningful signal and of noise at the input be

$$\Phi_{n_{\text{u}}}(j\omega)=\frac{1}{\omega^2+a^2};\ \Phi_{\text{ns}_{\text{u}}}(j\omega)=b^2.$$

Moreover, the transfer function of the system is

$$G_d(s)=\frac{K}{s}.$$

Substituting s for $j\omega$, we obtain the frequency response of the system:

$$G_d(j\omega)=\frac{K}{j\omega}.$$

Inserting these data in (4. 56a), we find the optimum frequency response of the system:

$$G_{\text{opt}}(j\omega)=\frac{\frac{1}{\omega^2+a^2}}{\frac{1}{\omega^2+a^2}+b^2}\cdot\frac{K}{j\omega}.$$

Substituting s for $j\omega$ and also s^2 for ω^2, we obtain the optimum transfer function of the system:

$$G_{\text{opt}}(s)=\frac{\frac{1}{-s^2+a^2}}{\frac{1}{-s^2+a^2}+b^2}\cdot\frac{K}{s}=-\frac{K}{b^2}\cdot\frac{1}{s^2-\frac{1+a^2b^2}{b^2}}\cdot\frac{1}{s}.$$

At its poles $G_{opt}(s)$ assumes infinite values. Since $K \neq 0$ and $b \neq \infty$, the poles of the function are determined from the equation

$$s\left(s^2-\frac{1+a^2b^2}{b^2}\right)=0.$$

The function thus has three poles:

$$s_1=0;\ s_2=\frac{\sqrt{1+a^2b^2}}{b};\ s_3=-\frac{\sqrt{1+a^2b^2}}{b}.$$

Since one of these poles ($s_2>0$) falls in the right halfplane, the optimum transfer function $G_{opt}(s)$ is physically unrealizable, as the system will be unstable.

Observe that in (4.56b) the term

$$\frac{\Phi_{n_{11}}}{\Phi_{n_{11}}+\Phi_{ns_{11}}} \tag{4.57}$$

is always a function of ω^2. The optimum transfer function therefore always has conjugate poles (such as s_2 and s_3 in the preceding example). This term will, therefore, *always* render $G_{opt}(s)$ unstable.

The only particular case when the optimum function is stable is the case of noiseless input, when (4.57) is equal to 1. In this case, if $G_d(s)$ is stable, the optimum system is naturally stable.

In all other cases, we are forced to choose a stable transfer function which best approximates the optimum function. There is no general procedure for approximating the optimum transfer function of a system by a stable function. We shall therefore consider the solution of Bode and Shannon for the following example.

It is required to design a stable system for *extrapolation* of a random function which would ensure minimum mean square error. Assume the input to be noiseless.

This problem is mathematically stated as follows:

$$u_2(t)=u_1(t+\alpha), \tag{4.58}$$

where $u_2(t)$ = the output signal;
$u_1(t)$ = the input signal;
α = the lead (extrapolation) time of the output signal.

The transfer function of the system is obtained by the Laplace transformation of (4.48), and is equal, according to equations (4.9) and (4.7), to

$$G_d=\frac{U_2(s)}{U_1(s)}=\frac{\int\limits_{-\infty}^{\infty}u_1(t+\alpha)e^{-st}\,dt}{\int\limits_{-\infty}^{\infty}u_1(t)e^{-st}\,dt}.$$

We substitute $\tau-\alpha$ for t in the upper integral (thus $dt=d\tau$).

The transfer function then has the following form:

$$G_d(s)=\frac{\int\limits_{-\infty}^{\infty}u_1(\tau)e^{-s\tau}\,e^{sa}\,d\tau}{\int\limits_{-\infty}^{\infty}u_1(t)\,e^{-st}\,dt}=e^{sa}.$$

Since by assumption the input is noiseless, we have from (4.56b)

$$G_{\mathrm{opt}}(s)=G_d(s)=e^{sa}. \tag{4.59}$$

This function has a pole, $s_1=\infty$. Since this pole falls in the right half-plane, the transfer function of (4.59) is unstable. We therefore must find a realizable function $G(s)$ which would best approximate the optimum function $G_{\mathrm{opt}}(s)$.

We solve this problem as follows. First, the physical system whose transfer function we want to find is represented in two parts (Figure 4.17), $G_1(s)$ and $G_2(s)$.

The spectral density of the meaningful signal fed into the sytem can be represented as a sum of two components:

$$\Phi_{11}(s)=\Phi_{11}^{+}(s)\,\Phi_{11}^{-}(s), \tag{4.60}$$

where $\Phi_{11}^{+}(s)$ = the complex component of the spectral density whose poles are all in the left halfplane;

$\Phi_{11}^{-}(s)$ = the complex component whose poles are all in the right halfplane.

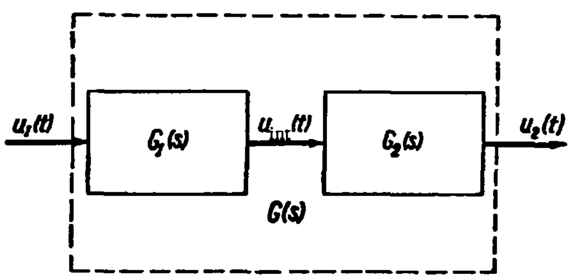

FIGURE 4.17. Determination of an approximation for $G_{\mathrm{opt}}(s)$.

For example, if

$$\Phi_{11}(j\omega)=\frac{1}{\omega^2+a^2}$$

or

$$\Phi_{11}(s)=-\frac{1}{s^2-a^2},$$

we take

$$\Phi_{11}^{+}(s)=\frac{1}{s+a};\ \Phi_{11}^{-}(s)=-\frac{1}{s-a}.$$

It is easy to see that always in this decomposition

$$\Phi_{11}^{-}(s)=\Phi_{11}^{+}(-s). \tag{4.60a}$$

The transfer function of the first part of the system is taken equal to

$$G_1(s)=\frac{1}{\Phi_{11}^{+}(s)}. \tag{4.61}$$

Since all the poles of $\Phi_{11}^{+}(s)$ are in the left halfplane, $G_1(s)$ is a physically stable function.

If now the transfer function of the second part of the system is taken equal to

$$G_2'(s)=\Phi_{11}^{+}(s), \tag{4.62a}$$

the transfer function of the entire system (see equations (4.61) and (4.62a)) is equal to 1:

$$G(s)=G_1(s)\ G_2'(s)=1.$$

In this case the input and output signals are equal.

$$u_2(t)=u_1(t).$$

Our object, however, is to obtain an output signal with a lead of α seconds. We shall therefore somewhat modify the transfer function of the second part of the system, setting it equal not to

$$G_2' \rightleftarrows g_2(t), \tag{4.62b}$$

where $g_2(t)$ is the impulse transfer function of the second part of the system defined as the inverse Laplace transform of $G_2'(s)$, but rather to

$$g_2(t+\alpha) \rightleftarrows G_2(s), \tag{4.63}$$

where $G_2(s)$ is the Laplace transform of $g_2(t+\alpha)$.

For this function to be stable, we introduce the restraint

$$g_2(t+\alpha)=0 \text{ when } t<0.$$

Note that in our system (Figure 4.17) the intermediate signal is white noise.

Indeed, the spectral power density of the intermediate signal $u_{\text{int}}(t)$ is equal, according to (4.43), to

$$\Phi_{\text{int}}(j\omega)=G_1(j\omega)\ G_1(-j\omega)\ \Phi_{11}(j\omega).$$

From (4.61)

$$G_1(s)=\frac{1}{\Phi_{11}^{+}(s)},$$

so that

$$G_1(j\omega)=\frac{1}{\Phi_{11}^{+}(j\omega)},$$

and from (4.60a)

$$G_1(-j\omega)=\frac{1}{\Phi_{11}^{-}(j\omega)}.$$

On the other hand, from (4.60)

$$\Phi_{11}(j\omega)=\Phi_{11}^{+}(j\omega)\,\Phi_{11}^{-}(j\omega),$$

whence

$$\Phi_{\text{int}}(j\omega)=1.$$

By having chosen the transfer function of the first part of the system according to (4.61), we have converted the signal $u_1(t)$ into white noise $U_{\text{int}}(t)$.

White noise can be considered, with some approximation, as a sequence of narrow random impulses coming at close intervals. Each of these impulses produces an output signal

$$[u_2(t)] \text{ from one impulse } =g_2(t).$$

The integration of these white-noise impulses gives the signal $u_2(t)$. The mean square value of the output function is determined as

$$\overline{u_2^2}=\int_0^{\infty} g_2^2(t)\,dt. \tag{4.64}$$

The signal $u_2(t)$ should lead $u_{\text{int}}(t)$ by α seconds. Hence, at $t=0$, the input signal fed into the second part of the system is equal to $u_{\text{int}}(0)$, and the output signal of the system should be $u_2(-\alpha)$. However, to satisfy the condition of stability, no output signal $u_2(t)$ should be produced during the time interval, $-\alpha$ to 0. The mean square error at the system output therefore amounts to neglecting the white-noise impulses between 0 and α seconds:

$$\overline{e^2}=\int_0^{\alpha} g_2^2(t)\,dt. \tag{4.65}$$

The relative mean square error at the output is therefore equal to

$$\frac{\overline{\varepsilon^2}}{\overline{u^2}}=\frac{\int\limits_0^{\alpha} g_2^2(t)dt}{\int\limits_0^{\infty} g_2^2(t)dt}. \tag{4.66}$$

Let us consider an example of determination of the transfer function for a stable system.

Consider the Laplace transform of the autocorrelation function of an input signal:

$$\Phi_{11}(s)=\frac{a^2}{(s^2-b^2)(s^2-c^2)}.$$

Design a system with a lead of α seconds.

The problem is solved as follows.

a) Following (4.60), $\Phi_{11}(s)$ is divided into two factors:

$$\Phi_{11}^{+}(s)=\frac{a}{(s+b)(s+c)};\ \Phi_{11}^{-}(s)=\frac{a}{(s-b)(s-c)}.$$

b) From tables, we determine the inverse function whose Laplace transform (see equations (4.62a) and (4.62b)) is $\Phi_{11}^{+}(s)$:

$$g_2(t)=\frac{a}{c-b}\left(e^{-bt}-e^{-ct}\right);\ (t>0). \tag{4.67}$$

c) The transfer function of the first block of the system is computed from (4.61).

$$G_1(s)=\frac{1}{\Phi_{11}^{+}(s)}=\frac{(s+b)(s+c)}{a}. \tag{4.68}$$

d) We determine from (4.67) the impulse transfer function of the second block of the system without considering its stability:

$$g_{2_{\text{opt}}}(t+\alpha)=\frac{a}{c-b}\left[e^{-b(t+\alpha)}-e^{-c(t+\alpha)}\right];\ (t>-\alpha). \tag{4.69}$$

e) The condition of stability requires that $g_{2\text{opt}}(t+\alpha)=0$ when $t<0$. Therefore, allowing for this condition, the transfer function (4.69) is represented as

$$g_2(t+\alpha)=\begin{cases}0;\ (t<0),\\ \dfrac{a}{c-b}\left[e^{-b(t+\alpha)}-e^{-c(t+\alpha)}\right];\ (t>0).\end{cases} \tag{4.70}$$

The transfer functions derived are shown in Figure 4.18. It follows from these curves that the function $g_2(t)$ (top graph) is stable (see 4.67). However a shift of α (bottom left) renders this function unstable (see 4.69). If we now

"cut off" from $g_{2\,opt}\,(t+\alpha)$ the part which protrudes into the range $t<0$ (bottom right), we obtain a stable function (4.70).

f) We now find from (4.63) the transfer function of the second block of the system, which is defined as the Laplace transform of $g_2(t+\alpha)$. To accomplish this, (4.70) is rewritten in a slightly different form:

$$g_2(t+\alpha)=\begin{cases}0; \ (t<0),\\ \dfrac{a}{c-b}\left[e^{-bt}e^{-b\alpha}-e^{-ct}e^{-c\alpha}\right]; \ (t>0).\end{cases}$$

Using Laplace-transform tables, we obtain

$$G_2(s)=\frac{a}{c-b}\left(\frac{e^{-b\alpha}}{s+b}-\frac{e^{-c\alpha}}{s+c}\right)=$$
$$=\frac{a\left[e^{-b\alpha}(s+c)-e^{-c\alpha}(s+b)\right]}{(c-b)(s+b)(s+c)}. \qquad (4.71)$$

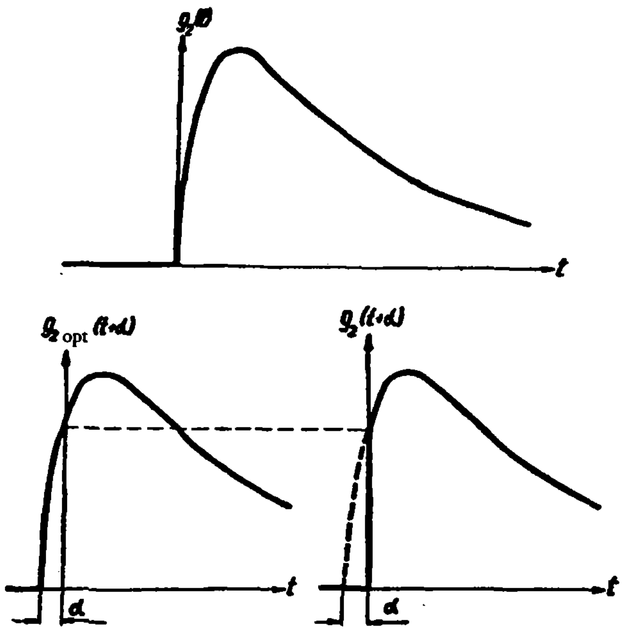

FIGURE 4.18. Transfer functions of the second block of the system.

g) We now determine the overall transfer function of the system:

$$G(s)=G_1(s)\ G_2(s).$$

Inserting $G_1(s)$ from (4.68) and $G_2(s)$ from (4.71), we obtain

$$G(s) = \frac{(s+b)(s+c)}{a} \cdot \frac{a\left[e^{-b\alpha}(s+c) - e^{-c\alpha}(s+b)\right]}{(c-b)(s+b)(s+c)} =$$
$$= \frac{e^{-b\alpha}(s+c) - e^{-c\alpha}(s+b)}{c-b}.$$

h) The mean square error of the lead output signal is equal, according to equations (4.65) and (4.67), to

$$e^2 = \int_0^\alpha g_2^2(t)\,dt = \int_0^\alpha \left[\frac{a}{c-b}\left(e^{-bt} - e^{-ct}\right)\right]^2 dt =$$
$$= \left(\frac{a}{c-b}\right)^2 \int_0^\alpha \left(e^{-2bt} - 2e^{-(b+c)t} + e^{-2ct}\right) dt =$$
$$= \left(\frac{a}{c-b}\right)^2 \left[\frac{1}{2b}\left(1 - e^{-2b\alpha}\right) - \frac{2}{b+c}\left(1 - e^{-(b+c)\alpha}\right) + \frac{1}{2c}\left(1 - e^{-2c\alpha}\right)\right].$$

i) The mean square value of the output function is determined from (4.64) Therefore, substituting ∞ for α in the preceding expression, we obtain

$$u_2^2 = \int_0^\infty g_2^2(t)\,dt = \left(\frac{a}{c-b}\right)^2 \left(\frac{1}{2b} - \frac{2}{b+c} + \frac{1}{2c}\right).$$

k) Finally, the relative error (4.66) is equal to

$$\frac{e^2}{u_2^2} = \frac{\frac{1}{2b}\left(1 - e^{-2b\alpha}\right) - \frac{2}{b+c}\left(1 - e^{-(b+c)\alpha}\right) + \frac{1}{2c}\left(1 - e^{-2c\alpha}\right)}{\frac{1}{2b} - \frac{2}{b+c} + \frac{1}{2c}}.$$

Chapter V

PRINCIPLES OF THE THEORY OF NONLINEAR AUTOMATIC-CONTROL SYSTEMS

In the previous chapters, only linear automatic-control systems were considered. Cybernetic systems, however, are nonlinear.

The nonlinearities arising in a system can be divided into two groups: nonlinearities inherent in the system elements, and nonlinearities introduced for the purpose of improving the system's response. In the first group we have saturation (Figure 5.1). backlash, temperature dependence, etc. In the second group we have nonlinear elements and nonlinear feedback.

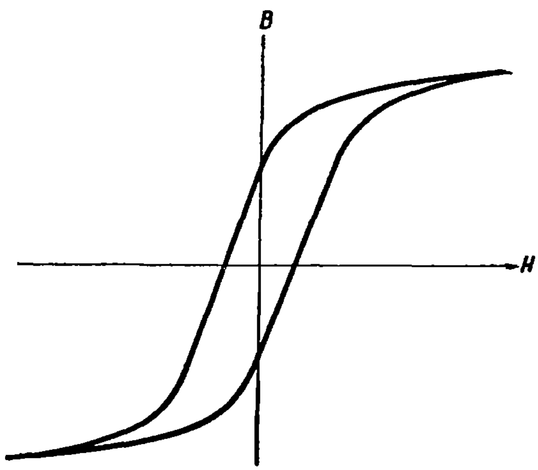

FIGURE 5.1. Hysteresis curve of ferromagnetic elements.

For example, Figure 5.2 shows the characteristic curve of a system hunting the extremum $(u_{2m};\ u_{1m})$. It can be seen from this characteristic that nonlinear elements have been introduced into the system to enable automatic control near these points (this problem will be discussed in more detail in the next chapter).

A non-linear system is essentially different from a linear system. The principle of superposition does not apply and the signal spectrum at the

output of a nonlinear system is different from the input signal spectrum. The stability of a nonlinear system is determined not only by the input signal, but also by the initial conditions. The transfer function of a nonlinear system depends on the input signal. It is therefore very difficult to calculate the response of nonlinear control systems. This involves the solution of complicated high-order, nonlinear differential equations, and a general analytical solution is possible only in a few particular cases with a varying degree of accuracy. At present numerous methods are available for the analysis of nonlinear systems. In what follows we shall dwell on those which are of fundamental interest.

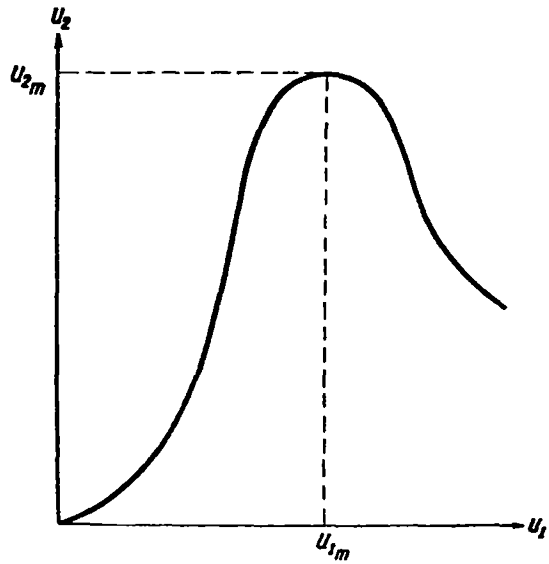

FIGURE 5. 2. Characteristic curve of an extremum system.

1. METHOD OF PIECEWISE LINEAR APPROXIMATION

This widespread method substitutes several straight segments (three in Figure 5.3) for the characteristic curve of a nonlinear element (the dashed curve in Figure 5.3). This method is highly convenient since it reduces the solution of a nonlinear problem to an approximate linear problem, and enables us to apply the techniques developed for linear systems. In some cases, however, the piecewise linear approximation method gives rise to very cumbersome mathematical expressions. Particularly complicated is the introduction of boundary conditions specifying the transition from one segment to another. This is the main factor limiting the applicability of the piecewise linear approximation.

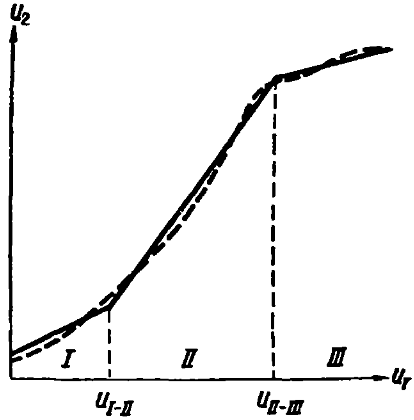

FIGURE 5.3. Piecewise linear approximation of the characteristic curve of a nonlinear element.

2. DESCRIBING-FUNCTION METHOD

The describing-function method replaces the nonlinear element of an automatic-control system by a linear unit which approximately describes its response.

This method makes the following assumptions.

1. There is a single nonlinear element in the system (if several nonlinear elements are introduced, they should be combined and considered as a single element).

2. *The response of the nonlinear element is time-independent.* This method therefore does not apply to nonlinearities introduced by heating of elements, drift of parameters, etc.

3. If a sinusoidal signal is applied to the input of the nonlinear element, its output is also a sinusoidal signal. In other words, only the fundamental component of the output signal is considered. This assumption is generally valid, because in most cases the higher harmonics of the signal have small (relative to the fundamental component) amplitudes and are thus negligible. The describing function approximating the response of a nonlinear element (Figure 5.4) can thus be represented in the form

$$N=\frac{D_1}{C}, \tag{5.1}$$

where N = the describing function;
C = the signal amplitude at the input of the nonlinear element;
D_1 = the amplitude of the fundamental component of the output signal.

It follows from this equation that the magnitude of the describing function depends on the input-signal amplitude fed into the nonlinear element. If, in addition, the nonlinear element contains an emf source, the describing

function depends on the frequency ω.

The introduction of a describing function thus makes possible the replacement of the nonlinear element by a linear unit with a definite gain and phase shift.

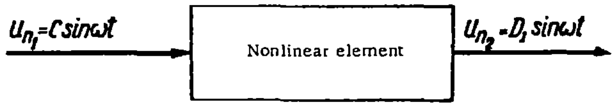

FIGURE 5.4. Nonlinear-element substitution diagram.

Let us consider an example of how a describing function is determined.

Take a closed-loop automatic-control system (Figure 5.5) consisting of three linear elements whose transfer functions are $G_1(s)$, $G_2(s)$, and $G_f(s)$, and a nonlinear element with no emf sources. A sinusoidal signal $u_1(t)$ of frequency ω is applied to the input.

We moreover assume that the nonlinear element has a saturation characteristic (Figure 5.6). The nonlinear element will therefore operate in two different ranges, depending on the amplitude of the input signal $u_{n_1}(t)$.

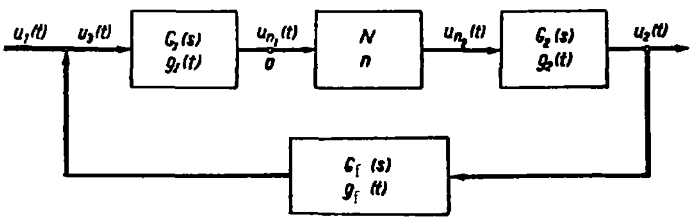

FIGURE 5.5. Circuit of a system with a nonlinear element.

If the signal amplitude $u_{n_1}(t)$ is not greater than u_a, the nonlinear element operates as a linear amplifier with a gain

$$k_n=\operatorname{tg}\alpha,$$

where α is the angle of slope of the linear portion of the characteristic.

If during part of the period $u_{n_1}(t)$ is greater than u_a, the nonlinear element operates as follows. During the first part of the period $(\beta\geqslant\omega t\geqslant 0$ (Figure 5.7), when $u_{n_1}(t)$ is smaller than u_a, the characteristic is linear. For this part of the period we may write

$$u_{n_2}(\vartheta)=k_nC\sin\vartheta;(\beta\geqslant\vartheta\geqslant 0), \tag{5.2}$$

where $k_n=\operatorname{tg}\alpha$ = the gain of the linear portion;

C = the amplitude of input signal fed into the nonlinear element;

ϑ = ωt.

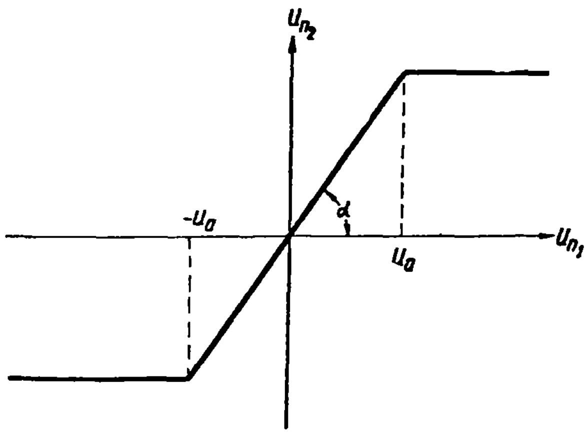

FIGURE 5. 6. Characteristic curve of the nonlinear element.

When the signal $u_{n_1}(t)$ exceeds u_a, the operating point of the nonlinear element shifts into the nonlinear saturation region (see Figure 5. 6). In this case the output signal of the nonlinear element remains constant:

$$u_{n_2}(\vartheta)=k_nC\sin\beta;\quad (\pi-\beta\geqslant\vartheta\geqslant\beta). \tag{5.3}$$

We thus see from equations (5. 2) and (5. 3) that for the first quarter period we may write

$$u_{n_2}(\vartheta)=\begin{cases} k_nC\sin\vartheta; & (\beta\geqslant\vartheta\geqslant 0),\\ k_nC\sin\beta; & \left(\frac{\pi}{2}\geqslant\vartheta\geqslant\beta\right). \end{cases} \tag{5.4}$$

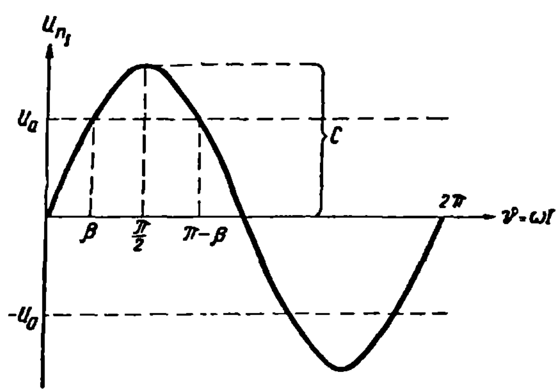

FIGURE 5. 7. The operation of a nonlinear element.

This signal is nonsinusoidal. To determine the describing function of the nonlinear element we must find the amplitude of the fundamental component of the output signal. It follows from (4.33) that this amplitude ($T=2\pi$) is

$$D_1=\frac{1}{\pi}\int_0^{2\pi} u_{n_2}(\vartheta)\sin\vartheta d\vartheta.$$

Changing the limits of integration (to integrate from 0 to $T/4$), we obtain

$$D_1=\frac{4}{\pi}\int_0^{\frac{\pi}{2}} u_{n_2}(\vartheta)\sin\vartheta d\vartheta.$$

Inserting $u_{n_2}(\vartheta)$ from (5.4) we find

$$D_1=\frac{4k_nC}{\pi}\left[\int_0^{\beta}\sin^2\vartheta d\vartheta+\sin\beta\int_{\beta}^{\frac{\pi}{2}}\sin\vartheta d\vartheta\right]=$$

$$=\frac{4k_nC}{\pi}\left[\left(\frac{\vartheta}{2}-\frac{1}{4}\sin 2\vartheta\right)_0^{\beta}+\sin\beta(-\cos\vartheta)_{\beta}^{\frac{\pi}{2}}\right]=$$

$$=\frac{4k_nC}{\pi}\left(\frac{\beta}{2}-\frac{1}{4}\sin 2\beta+\sin\beta\cos\beta\right).$$

Since

$$\sin 2\beta=2\sin\beta\cos\beta,$$

we have

$$D_1=\frac{2k_nC}{\pi}(\beta+\sin\beta\cos\beta). \quad (5.5)$$

The amplitude of the fundamental component at the output of the nonlinear element thus depends on the input amplitude of the signal (C) and the saturation voltage u_a.

Having determined the amplitude of the fundamental component (D_1) of the nonlinear element, we use (5.1) to construct the describing function:

$$N=\frac{2k_n}{\pi}(\beta+\sin\beta\cos\beta). \quad (5.6)$$

This function for $k_n=1$ is shown in Figure 5.8.

The parameter β determines, for a given frequency ω, the time (Figure 5.7) when $u_{n_1}=C\sin\vartheta$ equals u_a. This parameter is computed from the condition

$$C\sin\beta=u_a,$$

whence

$$\beta = \arcsin \frac{u_c}{C}. \tag{5.7}$$

Let us now evaluate the error introduced by neglecting the higher harmonics of the output-signal spectrum.

Since $u_{n_2}(\vartheta)$ in (5.4) is a curve symmetrical about the horizontal axis, it has no zero order d-c component and no even harmonics. Moreover, since this curve is symmetrical about the vertical axis passing through the point $\frac{\pi}{2}$ it is free from cosine functions. Therefore from (4.33) we have

$$u_{n_2}(t) = D_1 \sin \omega t + D_3 \sin 3\omega t + D_5 \sin 5\omega t + \ldots$$

and since $\omega t = \vartheta$, we obtain

$$u_{n_2}(t) = D_1 \sin \vartheta + D_3 \sin 3\vartheta + D_5 \sin 5\vartheta + \ldots$$

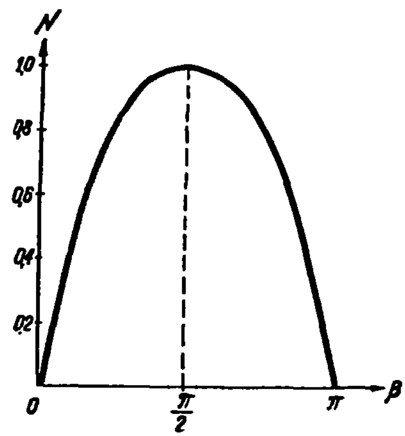

FIGURE 5.8. Describing function.

In order to estimate the error introduced when the third, fifth, etc. harmonics are dropped, we shall determine the amplitude of the third (the highest after the fundamental) harmonic. We have from (4.33) (integrating over a quarter period):

$$D_3 = \frac{4}{\pi} \int_0^{\frac{\pi}{2}} u_{n_2}(\vartheta) \sin 3\vartheta d\vartheta.$$

Inserting $u_{n_2}(\vartheta)$ from (5.4) and integrating we obtain

$$D_3 = \frac{4k_nC}{\pi}\left(\frac{\sin 2\beta}{4} - \frac{\sin 4\beta}{8} + \frac{\sin\beta\cos 3\beta}{3}\right)$$

or after simple manipulations

$$D_3 = \frac{4k_nC}{3\pi}\sin\beta\cos^3\beta.$$

Taking the ratio of the amplitudes of the third and fundamental harmonics

$$\frac{D_3}{D_1} = \frac{2}{3}\cdot\frac{\sin\beta\cos^3\beta}{\beta+\sin\beta\cos\beta}.$$

This ratio indirectly characterizes the accuracy of the analysis based on the describing function. If β is inserted from (5.7), we obtain the plot shown in Figure 5.9.

Let us now consider how the describing function can be used to analyze the stability of a nonlinear automatic-control system. The analysis will be carried out for the particular system shown in Figure 5.5.

This system is broken at point a, giving an open-loop system with the transfer function

$$G_p(s) = N\,G_2(s)\,G_f(s)\,G_1(s). \tag{5.8}$$

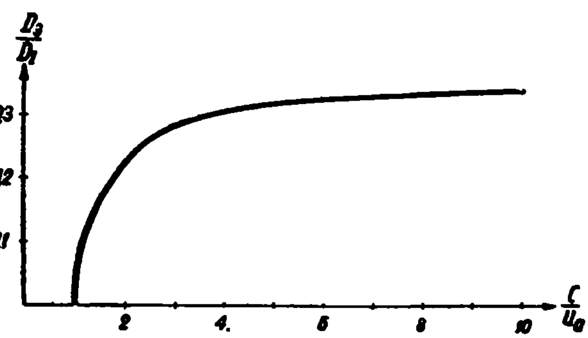

FIGURE 5.9. Amplitude ratio of the third and fundamental harmonics.

The stability is analyzed from the phase-amplitude response of the open-loop system using the Nyquist criterion. According to this criterion, a closed-loop automatic-control system is stable when the phase-amplitude response of the open-loop system $G_p(j\omega)$ does not contain the point $(-1;\ +j0)$. Since the introduction of the describing function N is equivalent to the substitution of a linear for a nonlinear element, the Nyquist criterion applies to the system (Figure 5.5) with the nonlinear element (Figure 5.6).

The limits of stability are determined by the equation (the phase-amplitude response passes through the point $(-1; +j0)$)

$$G_p(j\omega) = -1. \tag{5.9}$$

According to (5.8), condition (5.9) can be written as

$$-N = \frac{1}{G_1(j\omega)\,G_f\,(j\omega)\,G_2(j\omega)}. \tag{5.10}$$

We further plot in the complex plane $(+;+j)$ (Figure 5.10) the two functions described by equations (5.6) and (5.10):

$$-N = f(\beta)$$

$$\frac{1}{G_1(j\omega)\,G_f\,(j\omega)\,G_2(j\omega)} = \varphi(\omega).$$

The real part of the function $\varphi(\omega)$, and $f(\beta)$ are laid off on the abscissa of the complex plane, and the imaginary part of $\varphi(\omega)$ is plotted along the ordinate.

It follows from (5.10) that the limits of stability are determined by the intersection points of the two curves. Three cases are possible.

First case (Figure 5.10,a): $f(\beta)$ and $\varphi(\omega)$ do not intersect. The system is therefore stable for any β, and thus for any input signal $u_{n_1}(t)$ fed into the nonlinear element.

Second case (Figure 5.10,b): the system reaches the limit of its stability when β attains its maximum.

Third case (Figure 5.10,c): the system is stable when

$$\beta < \beta_a \text{ and } \beta > \beta_b$$

and unstable when

$$\beta_b > \beta > \beta_a.$$

The limiting values β_a and β_b are determined by the intersection points of the curves $f(\beta)$ and $\varphi(\omega)$.

We have previously considered a procedure for computing the describing function of an open-loop system. We shall now consider the determination of the describing function for a closed-loop system.

The difficulty in analyzing a closed-loop automatic-control system (Figure 5.5) is due to the fact that its describing function N depends, according to equations (5.6) and (5.7), on the amplitude C of the signal $u_{n_1}(t)$ fed into the nonlinear element. However, because of feedback, this amplitude depends, in its turn, on the transfer functions of the system elements and in particular on N. Moreover, in some cases, the nonlinearity of one of the elements may render the transfer function many-valued at some frequency (ω) of the input signal.

At present the method for determining the describing function of a nonlinear element operating in a closed-loop system has been developed only

for the case of a sinusoidal input with constant amplitude E and frequency ω:

$$u_1(t) = E \sin \omega t.$$

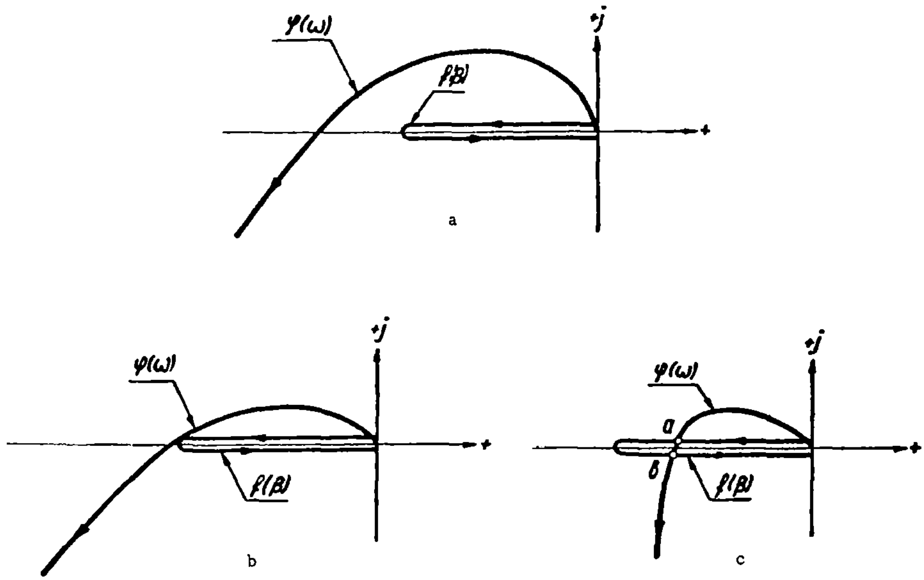

FIGURE 5.10. Determination of system stability.

Let us now consider how the describing function can be determined for the closed-loop system (Figure 5.5) containing a nonlinear element with a saturation characteristic (Figure 5.6).

The transfer functions of the linear elements are

$$G_1(s) = k;\ G_2(s) = \frac{k_2}{s+2};\ G_f(s) = \frac{k_f}{s}.$$

It follows from Figure 5.5 that the ratio between the Laplace transforms of the signals at the output of the system and the input of the elements $G_1(s)$ is equal to

$$\frac{U_2(s)}{U_3(s)} = NG_1(s)G_2(s). \tag{5.11}$$

Moreover,

$$\frac{U_{n_1}(s)}{U_3(s)} = G_1(s). \tag{5.12}$$

The ratio between the Laplace transforms of the input and output signals

of the closed system is determined from (4.14) as

$$G'(s)=\frac{U_2(s)}{U_1(s)}=\frac{G(s)}{1+G(s)G_f(s)},$$

where $G(s)=NG_1(s)G_2(s)$ is the transfer function of the open-loop system (no feedback).

Inserting $G(s)$ into this expression, we obtain

$$\frac{U_2(s)}{U_1(s)}=\frac{NG_1(s)G_2(s)}{1+NG_1(s)G_2(s)G_f(s)}. \tag{5.13}$$

Solving equations (5.11), (5.12) and (5.13) simultaneously, we find the ratio required in further computations:

$$\frac{U_{n_1}(s)}{U_1(s)}=\frac{G_1(s)}{1+NG_1(s)G_2(s)G_f(s)}. \tag{5.14}$$

Inserting $G_1(s)$, $G_2(s)$, and $G_f(s)$ into this equation, we obtain

$$\frac{U_{n_1}(s)}{U_1(s)}=\frac{k_1}{1+\dfrac{k_1k_2k_fN}{s(s+2)}}=\frac{k_1s(s+2)}{kN+s(s+2)},$$

where $k=k_1k_2k_f$.

Substituting $j\omega$ for s, we have

$$\frac{U_{n_1}(j\omega)}{U_1(j\omega)}=\frac{k_1j\omega(j\omega+2)}{kN+j\omega(j\omega+2)}.$$

We rewrite this equation in a somewhat different form:

$$\frac{U_{n_1}(j\omega)}{U_1(j\omega)}=\frac{k_1(-\omega^2+j2\omega)}{kN-\omega^2+j2\omega}=k_1\frac{\omega^2(4+\omega^2-kN)+j2\omega kN}{(kN-\omega^2)^2+4\omega^2}.$$

Applying (4.11), we again rewrite this equation as

$$\frac{U_{n_1}(j\omega)}{U_1(j\omega)}=A(\omega)\,e^{j\upsilon(\omega)},$$

where

$$A(\omega)=\sqrt{B_1^2(\omega)+B_2^2(\omega)};$$

$$\operatorname{tg}\upsilon=\frac{B_2(\omega)}{B_1(\omega)}.$$

In this case, the real component is

$$B_1(\omega)=k_1\frac{\omega^2(4+\omega^2-kN)}{(kN-\omega^2)^2+4\omega^2},$$

and the imaginary component is

$$B_2(\omega)=k_1\frac{2\omega kN}{(kN-\omega^2)^2+4\omega^2}.$$

The amplitude ratio of the signals $u_{n_2}(t)$ and $u_1(t)$ is therefore

$$A(\omega)=\frac{C}{E}=k_1\frac{\sqrt{\omega^4(4+\omega^2-kN)^2+(2\omega kN)^2}}{(kN-\omega^2)^2+4\omega^2}.$$

After further manipulation

$$C=k_1E\sqrt{\frac{\omega^2(\omega^2+4)}{(kN-\omega^2)^2+4\omega^2}}. \qquad (5.15)$$

Here k, k_1, E, and ω are known, and C and N are unknown.

On the other hand, the describing function N is related by equations (5.6) and (5.7) with the amplitude C of the input signal fed into the nonlinear element; we have

$$N=\frac{2k_n}{\pi}\left[\arcsin\frac{u_a}{C}+\frac{u_a}{C}\sqrt{1-\left(\frac{u_a}{C}\right)^2}\right], \qquad (5.16)$$

where $k_n=\operatorname{tg}\alpha$ and u_a are known parameters of the nonlinear element (Figure 5.6).

If equations (5.15) and (5.16) are now solved simultaneously, we obtain the describing function N.

3. PHASE-PLANE METHOD

Analysis of automatic-control systems by the describing-function method essentially reduces the problem to the substitution of the given nonlinear system by an equivalent linear system. In distinction from this method, the phase-plane method provides a technique for the direct analysis of nonlinear systems.

The phase-plane method amounts to the following. Take an automatic-control system described by a second-order nonlinear differential equation. Its state can be specified by two parameters in a two-dimensional plane, i.e., on the so-called *phase plane*. An analysis of the phenomena occurring on the phase plane yields the essential parameters of the system Therefore, the phase-plane method can be applied only to systems described by differential equations of order no higher than 2. Moreover, the initial values of the parameters represented on this plane must be given.

Before proceeding with the analysis of nonlinear systems, consider a linear system described by a second-order differential equation

$$\frac{d^2u}{dt^2}+a_1\frac{du}{dt}+a_0u=0. \qquad (5.17)$$

The initial conditions of the system at zero time are

$$\left(\frac{du}{dt}\right)_{t=0}=u_0';\ u_{t=0}=u_0.$$

First, let us consider a particular case of (5.17) when $a_1=0$. Passing from the variable t to the complex-frequency domain (s), we write

$$s^2U(s)+a_0U(s)=su_0+u_0',$$

whence we obtain the Laplace transform of the signal

$$U(s)=\frac{su_0+u_0'}{s^2+a_0}=u_0\frac{s}{s^2+a_0}+u_0'\frac{1}{s^2+a_0}.$$

From Laplace transform tables we obtain the inverse function

$$u=u_0\cos\sqrt{a_0}\,t+\frac{u_0'}{\sqrt{a_0}}\sin\sqrt{a_0}\,t.$$

Set

$$C\sin\varphi=u_0;\ C\cos\varphi=\frac{u_0'}{\sqrt{a_0}}. \tag{5.18}$$

Then

$$u=C\sin\left(\sqrt{a_0}\,t+\varphi\right), \tag{5.19}$$

where from (5.18)

$$C=\sqrt{u_0^2+\frac{(u_0')^2}{a_0}};$$

$$\varphi=\operatorname{arctg}\frac{\sqrt{a_0}\,u_0}{u_0'}. \tag{5.19a}$$

It follows from (5.19) that

$$\frac{du}{dt}=\sqrt{a_0}\,C\cos\left(\sqrt{a_0}\,t+\varphi\right). \tag{5.20}$$

We rewrite (5.19) in the following form:

$$u=C\sqrt{1-\cos^2\left(\sqrt{a_0}\,t+\varphi\right)}.$$

Inserting $\cos\left(\sqrt{a_0}\,t+\varphi\right)$ from (5.20) we obtain

$$\left(\frac{u}{C}\right)^2+\left(\frac{1}{\sqrt{a_0}C}\cdot\frac{du}{dt}\right)^2=1. \tag{5.21}$$

If now the variable u is plotted along the horizontal axis (Figure 5.11), and $\frac{du}{dt}$ along the vertical axis, (5.21) is represented by an ellipse with the semi-axes

$$d_1 = C; \ d_2 = \sqrt{a_0}\,C.$$

The coefficient C defined in (5.19a) changes in accordance with the initial conditions (u_0 and u_0'), tracing a family of ellipses on the plane in Figure 5.11. This family is called the phase portrait of the automatic-control system, defined by (5.17) with $a_1 = 0$.

The phase portrait thus specifies the dependence of the first derivative of the signal on the amplitude of this signal. We see in Figure 5.11 that time is not explicit in the phase portrait. The arrows therefore indicate the direction of displacement of the operating point of the system in time (the direction of the phase trajectory).

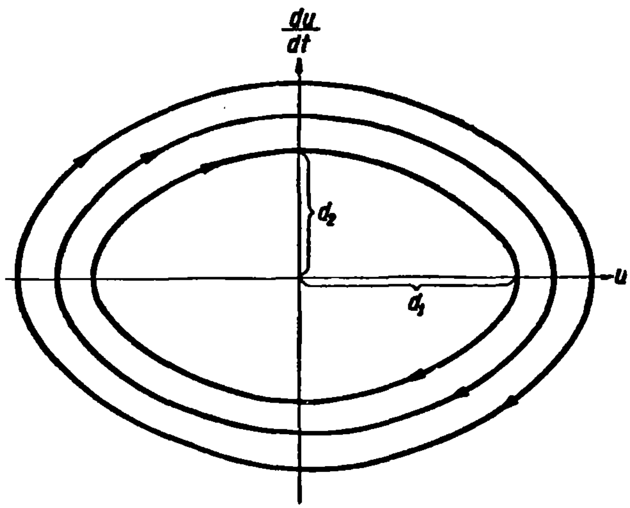

FIGURE 5.11. Phase portrait of a system with stable oscillations.

The points of the phase plane where

$$\frac{d^2u}{dt^2} = 0 \text{ and } \frac{du}{dt} = 0 \tag{5.22}$$

are called singular. Since the derivatives of u and $\frac{du}{dt}$ vanish at these points, the singular points correspond to a state of rest of the system. If a system is stable, its phase trajectory approaches a singular point.

For the case represented by (5.19), the singular point is determined as

follows. First, from (5.22) it follows that

$$\frac{du}{dt}=0.$$

The singular point thus falls on the horizontal axis.

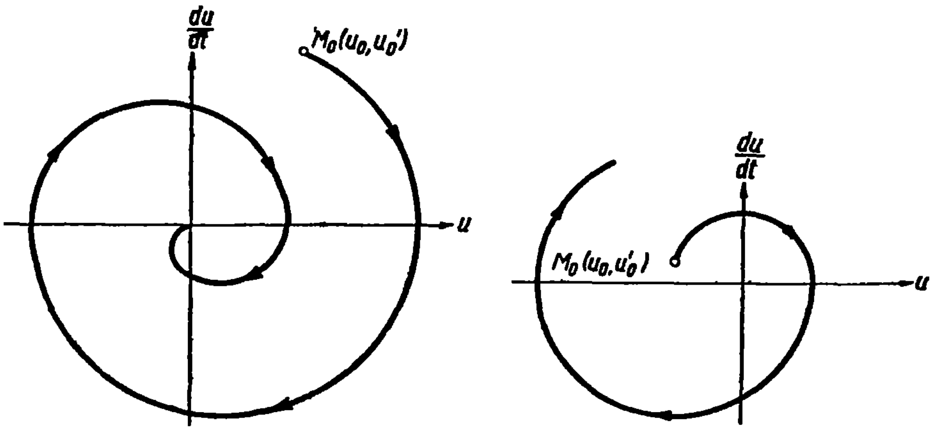

FIGURE 5.12. Phase trajectories of a stable (a) and an unstable (b) system.

The second derivative of the signal is equal from (5.19) to

$$\frac{d^2u}{dt^2}=-a_0C\sin(\sqrt{a_0}\,t+\varphi)=-a_0u=0.$$

Therefore, since $a_0 \neq 0$, the second coordinate of the singular point on the phase plane is equal to

$$u=0.$$

In this case the singular point is the origin. In our example (Figure 5.11) the working point moving along the phase trajectory oscillates about the origin. The ellipse therefore represents stable oscillations of the system. The amplitude of these oscillations is determined by the initial conditions u_0 and u_0'.

When the coefficient a_1 in (5.17) is not zero, the ellipse degenerates into a spiral. If the system is stable, its operating point (Figure 5.12,a) moves along the spiral to a singular point (here, the origin). If, however, the system is unstable, the operating point (Figure 5.12,b) moves away from the singular point tracing an unwinding spiral. Plotting the spirals for various initial conditions, we obtain the phase portrait of the system. The phase portrait of a stable system having the origin as its singular point is shown in Figure 5.13.

We have previously observed that time is not explicit in the phase portrait. It is nevertheless possible to determine from the phase portrait

the time of motion of the operating point from one position on the phase trajectory to another. The time of motion of the operating unit from one position (A) to another (B) is determined as follows:

$$dt = \frac{1}{\frac{du}{dt}} du.$$

Integrating this equation, we obtain

$$t = \int_{u_A}^{u_B} \frac{1}{\frac{du}{dt}} du.$$

Hence, plotting the curve $\frac{1}{\frac{du}{dt}} = f(u)$, we obtain the time from the area between this curve and the horizontal axis.

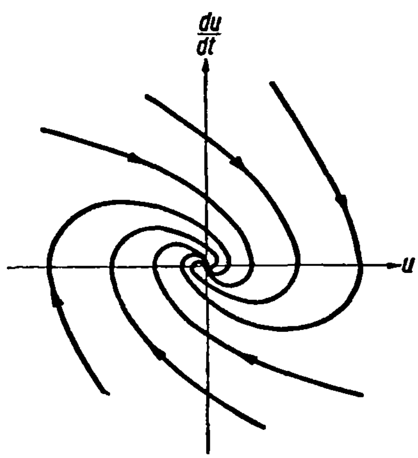

FIGURE 5.13. Phase portrait of a stable linear system.

The analysis of nonlinear automatic-control systems on the phase plane is carried out exactly as for linear systems. No general analytical solution of the nonlinear equations is attempted, since in most cases this is simply imposssible.

To obtain the phase portrait of a nonlinear system, we can use the circuit shown in Figure 5.15. In this circuit the signal (u) of the system to be analyzed is fed to the horizontal-sweep of an oscilloscope. Simultaneously, a differential unit feeds the rate of change of this signal $\frac{du}{dt}$ to the vertical sweep. The oscilloscope therefore traces the phase trajectory of the system. Varying the initial conditions, we may reproduce the phase portrait of the system.

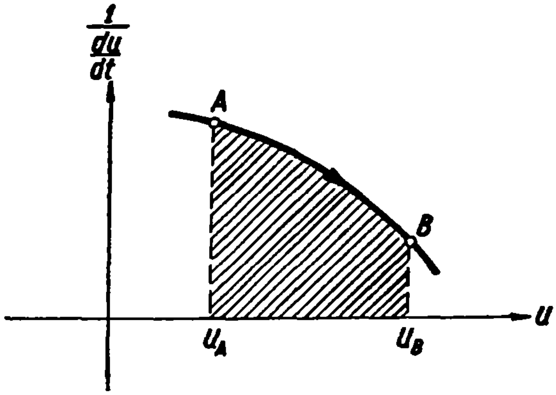

FIGURE 5. 14. Determination of time of motion of an operating point.

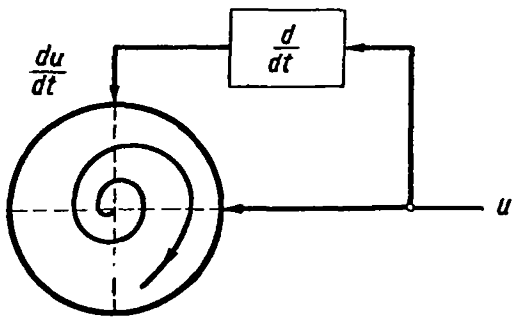

FIGURE 5.15. Circuit for producing the phase portrait of a nonlinear system.

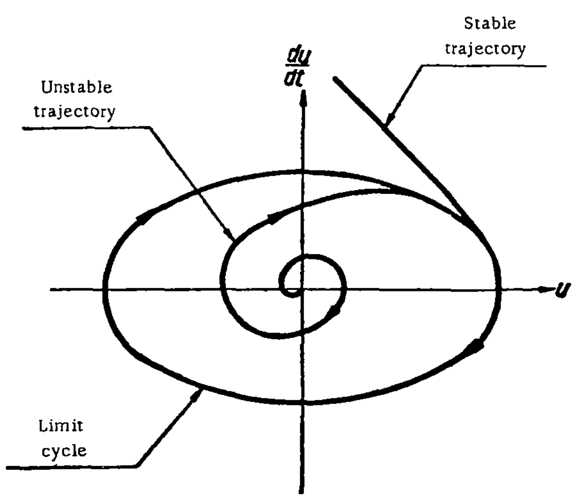

FIGURE 5.16. Phase trajectories of a nonlinear system.

A certain important feature of phase portraits of nonlinear systems must be kept in mind. The phase trajectories of linear systems were either stable or unstable, whereas in the case of nonlinear systems we may obtain trajectories some of which are stable and others are unstable.

As an example, Figure 5.16 shows phase trajectories of a nonlinear system. We see from this figure that one trajectory is stable and the other is unstable. Between these two we have a trajectory corresponding to a mode of stable oscillations. This trajectory is called the limit cycle.

Chapter VI

ADAPTIVE CONTROL SYSTEMS

1. GENERAL CONCEPTS

Until recently only ordinary automatic-control systems were used in the national economy. These systems controlled some specified parameter of the controlled object, by measuring its deviation from a nominal reference value and taking appropriate action to eliminate the arising error.

One of the simplest ordinary systems is shown in Figure 6.1. This system comprises a measuring device (a nonlinear bridge) and a magnetic amplifier (M). The controlled object is a synchronous generator whose voltage (u_g) must be maintained constant.

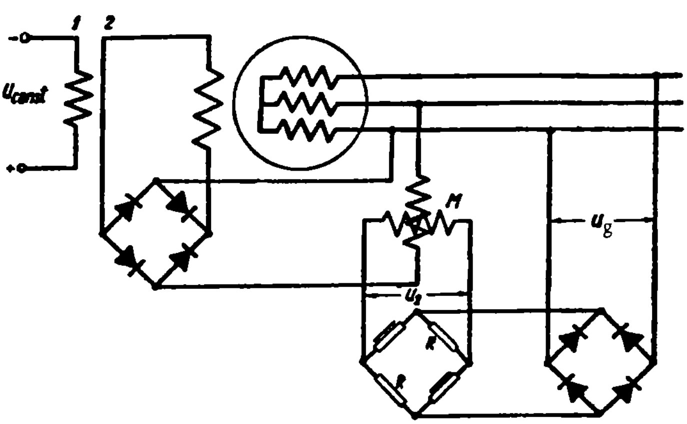

FIGURE 6.1. Synchronous generator voltage regulation.

The measuring bridge consists of two linear resistors (R) and two nonlinear elements (zener diodes, thermistor, etc.) whose resistances vary with current. The characteristic of the measuring device is represented by the curve in Figure 6.2. The operating portion (solid line) is described by the equation

$$u_1 = k(u_{gn} - u_g), \tag{6.1}$$

where $k = \mathrm{tg}\,\alpha$ = the gain;

u_{gn} = the nominal generator voltage.

The system functions as follows. When the generated voltage is at its nominal value, the excitation current is produced by field winding (*1*) which is supplied from a d-c voltage source u_{const}. According to (6.1), when the nominal voltage is generated, the output voltage signal from the bridge is zero. The magnetic amplifier is therefore cut off and no current flows through field winding (*2*) of the generator (idling current is neglected).

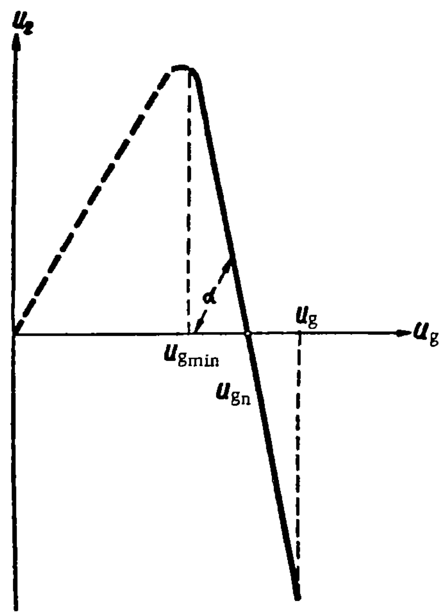

FIGURE 6.2. Characteristic curve of the measuring device.

If now for some reason (e.g., a load is connected to the circuit) the generator voltage drops, a negative voltage signal is produced at the output of the bridge (see equation (6.1)). This signal is proportional to the deviation of the generated voltage from its nominal value. The impedance of the magnetic amplifier therefore decreases and current flows through winding (*2*). This current reduces the deviation of the generated voltage from its nominal value to a permissible value.

The development of engineering cybernetics has contributed to the rapid advent of a new type of automatic-control systems, the so-called adaptive (or self-adjusting) control systems. Ordinary control systems have a rigidly defined regulation specification (e.g., (6.1)) to ensure the constancy of a certain parameter, whereas adaptive systems are intended for solving much more complicated problems which require adaptation (or adjustment) of the regulation specification. In most cases,

adaptive systems are used to ensure extremum (maximum or minimum) values of various parameters of the controlled object.

For example, an automatic-control system used in chemical industries can be designed to regulate a particular process so that maximum output (of given product quality) is ensured with minimum consumption of raw material. These systems are known as optimum systems. The parameters which are optimized in automatic control are called extremal.

Optimum adaptive systems can be of two kinds: variable-structure systems and constant-structure systems. Since variable-structure systems have not been very much studied, we shall consider in what follows only constant-structure systems. The discussion will be limited to cases where the controlled parameter has a single extremum which depends on a single independent variable.

The simplest adaptive system can be designed by introducing an additional element into the ordinary automatic-control systems. This element (K) can be coupled either in the forward loop of the controlled member (Figure 6.3) or in the feedback loop. The element K is so designed that its parameters adapt themselves in accordance to the control signal emitted by the logical computing circuit. The computing circuit, in turn, measures the input and output parameter of the system, processes the information received, and controls the element K so that, regardless of any change in the object or in external conditions, the controlled parameter is extremized.

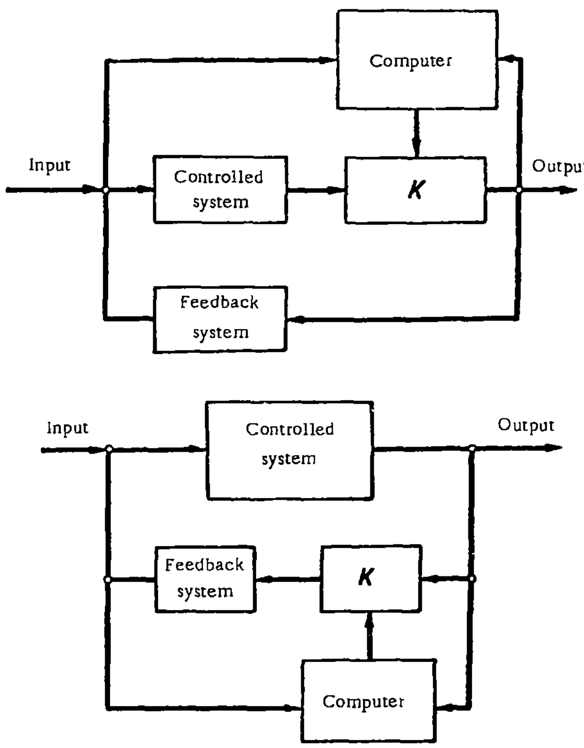

FIGURE 6.3. Simple circuits for adaptive control systems.

Adaptive systems are always very complicated logically and are invariably nonlinear. Their analysis and design is therefore a formidable problem.

Before proceeding with a discussion of adaptive control systems, we shall consider a system with a time-fixed extremum and a programmed control system. These systems are much simpler than adaptive control systems. We should therefore be well aware of the range of their application, so as to restrict the use of complex adaptive control systems to cases where the problem cannot be solved by these two simpler systems. Moreover, the analysis of cybernetic extremum systems will help in understanding the design and the operation of adaptive control systems.

2. FIXED-EXTREMUM SYSTEMS

We shall consider this type of automatic-control systems by the following example.

Consider a heavy-duty semiconductor rectifier, whose characteristic is shown in Figure 6.4. The current (i) through the rectifier is to be controlled for maximum efficiency (η_{max}) of the device.

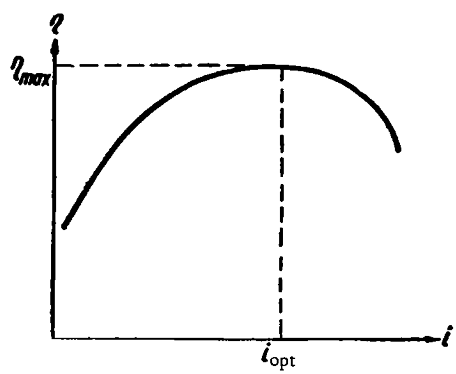

FIGURE 6.4. Characteristic curve of a semiconductor rectifier.

When the maximum efficiency is constant in time (being independent of heating and aging of the rectifier), we say that the controlled object has a time-fixed (or simply fixed) extremum. From Figure 6.4 it is easy to see that, to maintain maximum efficiency, the control system must be designed in accordance with the following regulation specification:

$$i_{opt} = \text{const.}$$

Hence, if an ordinary regulator receives a signal proportional to the rectifier current, this regulator will maintain constant optimum current; thus maximizing the efficiency.

A suitable system is shown in Figure 6.5. The rectifier bridge B and its load Z are coupled through a magnetic amplifier M to an a-c source (u_n). One control winding of the magnetic amplifier receives a constant voltage $u_{const} = Ri_{opt}$ and the other control winding receives a voltage proportional to the rectifier current $u = Ri$. Since the fluxes induced by these currents in the magnetic amplifier are in opposite directions, this circuit ensures that the rectifier operates at $i_{opt} = \text{const}$. Indeed, let the current i increase for some reason (e.g., an increase of input voltage u_n). The voltage applied to the second control winding of the amplifier increases. Consequently, the amplifier impedance increases and the current i drops approximately to its previous value.

Since the current at the amplifier output is constantly equal to i_{opt}, the efficiency of the rectifier is maximized (see Figure 6.4).

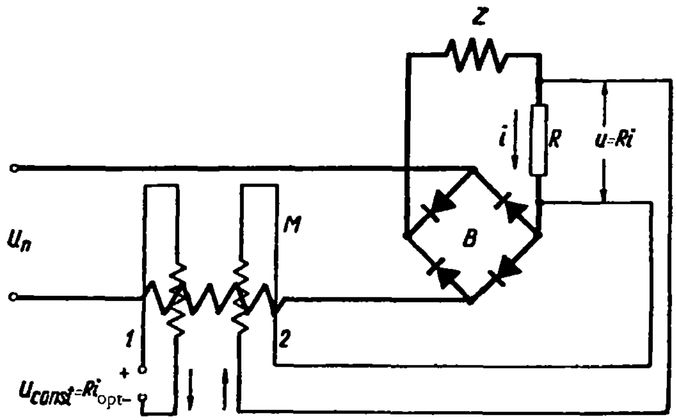

FIGURE 6.5. Circuit of a control system with fixed extremum.

3. PROGRAMMED SYSTEMS

These systems are used when we have a fairly accurate mathematical description of the relationship between the extremal parameter and the parameters affecting its variation; moreover, this relationship should not be unduly complicated.

Consider an example of a programmed control system. Our problem is to design an automatic-control system ensuring aircraft will cover the maximum possible distance. The flight range of aircraft depends on its fuel consumption, i.e., on the efficiency of its engines. Aircraft engines have a specified optimum operating condition where a minimum amount of fuel per km $\left(\frac{Q}{l}\right)$ is burned. Consequently, an optimum flight velocity (V_1) exists (Figure 6.6) for minimum fuel consumption.

Thus, except for one essential factor, we could consider aircraft as fixed-extremum systems, analogous in principle to the previous case of the semiconductor rectifier. A fixed velocity equal to V_1 would have then ensured maximum flight range.

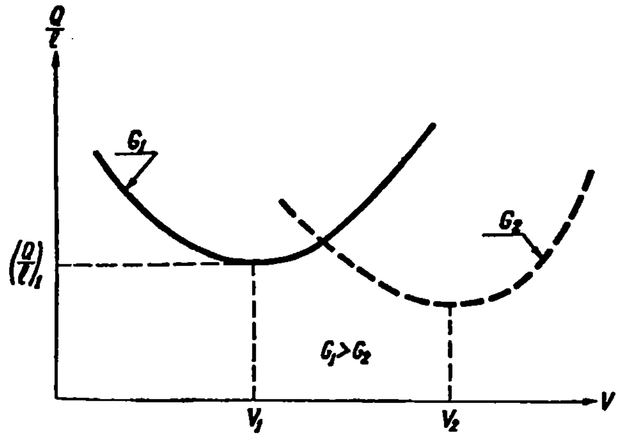

FIGURE 6.6. Specific fuel consumption as a function of flight velocity.

However, in long-distance flights the weight of the fuel constitutes a considerable portion of the aircraft weight. Therefore, as fuel is burnt, the weight G of the aircraft decreases. Since the optimum velocity changes as the aircraft becomes lighter (the dashed curve in Figure 6.6), the extremum system is disturbed. Hence, as the weight of aircraft decreases, the extremal velocity must be readjusted. This adjustment is determined as follows.

The extremal velocity of the aircraft as a function of its weight is

$$V_{ex} = f(G).$$

The variation of the aircraft weight in flight is

$$G = \varphi(t).$$

These two functions give the regulation specification for the flight velocity, ensuring the maximum range for the aircraft:

$$V_{ex} = f[\varphi(t)]. \tag{6.2}$$

The flight velocity can thus be controlled in accordance with relationship (6.2) by a programed control system. This control system consists of a programed device (magnetic or perforated tape) storing the function (6.2), and of a system for adjusting the aircraft controls. The accuracy of the automatic system of course depends on how equation (6.2) describes the true character of the changes in the flight velocity.

4. DERIVATIVE-SENSING SYSTEMS

In the preceding cases the control equations were known. Let us now consider some cases where these equations are unavailable.

A diesel generator operating in parallel with a power system, must be so controlled that the overall efficiency is maximum at all times.

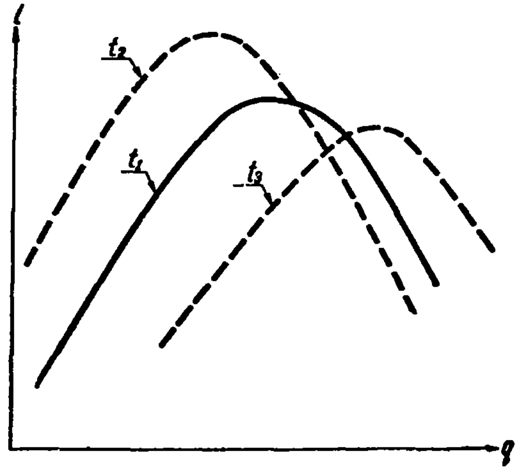

FIGURE 6. 7. Efficiency versus fuel-consumption characteristics of a diesel generator.

The previous examples of extremum control systems do not apply here. First, the equations describing the operating conditions of the diesel generator are unavailable and, second, changes in diesel temperature and in grid parameters preclude the choice of operating conditions which would remain optimum during the entire relevant time. The relationship between the efficiency of the diesel generator and the fuel (q) required is shown by the solid curve in Figure 6.7. No mathematical description for this curve is available. Moreover, this curve becomes distorted, and drifts in a rather arbitrary way during the time t (the dashed curves).

However, regardless of the actual function $\eta=f(q)$, the extremum efficiency of the diesel generator is determined by the relationship

$$\frac{d\eta}{dq}=0.$$

To ensure extremum control of the diesel generator we therefore compute its efficiency, differentiate it with respect to the quantity of fuel, and maintain the derivative at zero value at all times.

A circuit of the system providing this control is shown in Figure 6.8. The input of the systems consists of the output parameters of the generator (current, voltage, and $\cos\varphi$), and the input of the diesel (fuel). Multiplying the three generator parameters, we obtain the output power (W_{out}) of the diesel generator. Since the amount of fuel fed into the diesel (q) is proportional to the generated output power, the quotient W_{out}/q gives the efficiency. The rates of change of the efficiency $\left(\frac{d\eta}{dt}\right)$, and of the fuel consumption $\left(\frac{dq}{dt}\right)$

are computed. The ratio of these two parameters gives the derivative $\frac{d\eta}{dq}$. The signal $u = \frac{d\eta}{dq}$ is compared with zero by an ordinary automatic regulator.

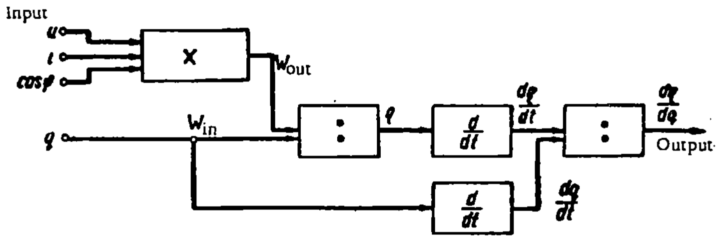

FIGURE 6.8. Circuit differentiating the efficiency with respect to the quantity of fuel.

It can be seen from Figure 6.8 that a derivative sensing system requires at least three units: two differentiations and one divider.

The derivative-sensing extremum system is fairly simple. However, differentiators, as we have observed in section 2-9,d, amplify high-frequency errors. This liability sharply limits the application of systems of this kind.

5. SIGN-SENSING SYSTEMS

In order to eliminate differentiation and their associated errors, units have been developed to react to the sign of the derivative, rather than to its magnitude. A system using these units for maximizing the efficiency of a diesel generator will now be considered.

The efficiency η varies with the quantity of fuel input q according to the curve shown in Figure 6.9. This curve drifts arbitrarily in time as a result of changes in the parameters of the diesel generator. We shall assume, that this drift is much slower than the response time of the control system. Moreover, we shall tentatively assume all the system components to be intertia-less. Our problem is then reduced to the design of a control system which would adjust itself to the extremum of a *static* (nondrifting) curve $\eta = f(q)$.

First, the derivative is replaced by a ratio of finite increments:

$$\frac{d\eta}{dq} = \frac{\Delta\eta}{\Delta q}.$$

An electric motor operates the fuel feed pump. If the efficiency is increasing, the diesel generator is operating in the left-hand part of the characteristic (Figure 6.9) moving toward the extremum point. When η has reached the extremum on its upward path, the efficiency starts decreasing. Therefore, if q increases, the efficiency η changes its trend from upward to downward

and the electric motor must be reversed. When q is decreasing, and the efficiency changes its trend from upward to downard, the motor must again be reversed.

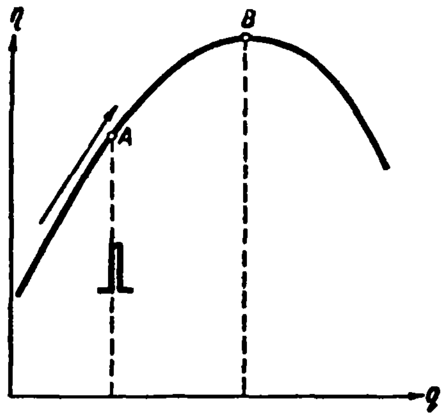

FIGURE 6.9. The characteristic curve of the controlled parameter.

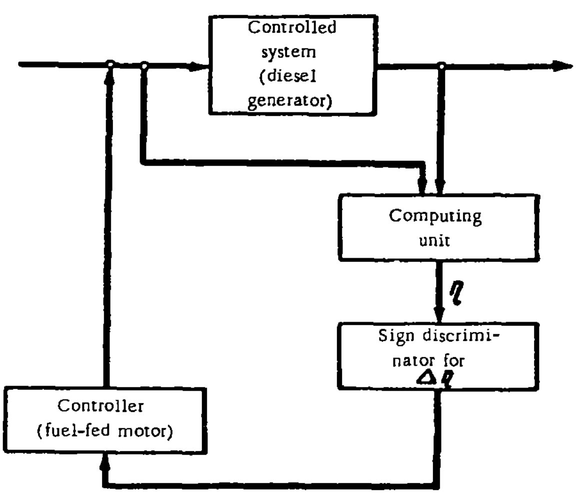

FIGURE 6.10. Sign-sensing system.

We thus see that in an extremum-hunting control system, the fuel supply must be reversed whenever the increment $\Delta\eta$ changes its sign. This logical proposition is the basis on which these sign-sensing systems operate.

A circuit of a sign-sensing adaptive control system is shown in Figure 6.10. In this circuit the computor receives signals from the input and the output of the controlled member and computes the efficiency η.

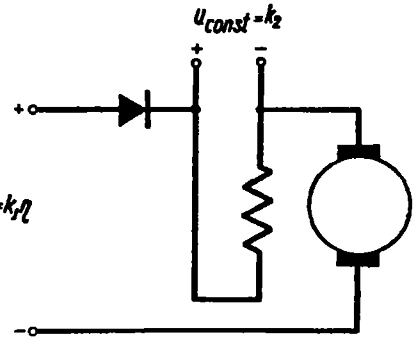

FIGURE 6.11. Determination of the sign of the increment.

Note that in adaptive control systems the output of the computing unit which determines the extremal parameter (in our case, η) is called the system output. The computing device is thus an inseparable part of the controlled member. The term "system output" will be used in this special sense in what follows.

The system output (Figure 6.10) is fed into a sign discriminator for determining the sign of $\Delta\eta$. Whenver the sign changes, the motor actuating the fuel feed pump is reversed.

A circuit for determining the sign of the increment $\Delta\eta$ is shown in Figure 6.11. Since in this circuit the d-c motor has constant (in direction and magnitude) excitation current, its speed is proportional to the armature voltage. Hence with $k_1\eta>k_2$, the motor speed is

$$n=k_3(k_1\eta-k_2)$$

where k_3 is a coefficient of proportion.

The speed-voltage characteristic is shown in Figure 6.12. The diode connected in the circuit (Figure 6.11) clips the superfluous negative portion of this characteristic (the dashed line).

A centrifugal switch on the rotor of this auxiliary motor makes contact whenever the motor changes over from acceleration to deceleration. The motor decelerates when the efficiency starts decreasing. The centrifugal switch can thus be used to provide the reversal signal for the fuel-feed motor, as shown in Figure 6.13. In this circuit the coils (C_1 and C_2) of two selectors reverse the polarity of the excitation winding and thus the direction of the fuel-feed motor whenever the contact of the centrifugal switch (S) is made.

Let us consider the operation of the system from the moment the fuel-feed motor is started and the supply of fuel (q) starts increasing (from zero). The efficiency η increases simultaneously (Figure 6.9).

When the extremum B is attained, the efficiency starts decreasing. The switch S (Figure 6.13) closes and the selector is rotated through one division, reversing the polarity of the current in the windings of the fuel-feed motor. The motor is reversed and the fuel supply (q) is gradually reduced. The efficiency increases again, but passing the extremum it starts decreasing.

The switch is again actuated and the fuel-feed motor is reversed. The quantity of fuel again increases producing a simultaneous increase in η. The system thus continuously h u n t s for the extremum. This oscillatory motion about the extremum is called h u n t i n g of the control system. During hunting, the system as if follows the receding motion of the extremum and takes appropriate logical measures for catching up with this point.

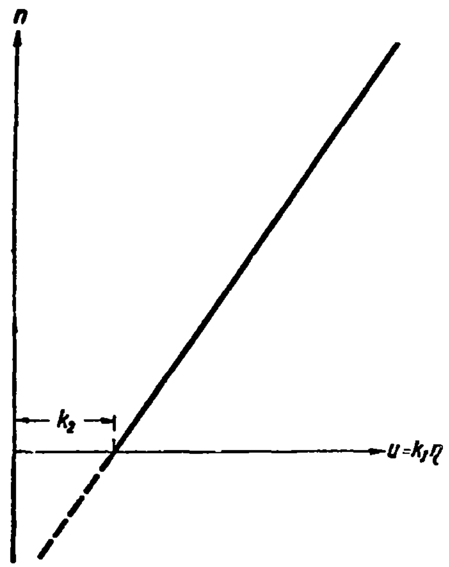

FIGURE 6.12. Speed-voltage characteristic.

There is a serious shortcoming in this system. For example, assume q is increasing and the system reaches point A (Figure 6.9). The extremum specifications do not require the reversal of q at this point, since η is increasing. Let now a transient pulse appear at the system output. The decay of this pulse (decrease of voltage) is interpreted by the circuit of Figure 6.10 as a decrease of η. Switch S thus makes contact and the fuel feed is reversed. Correspondingly, q will decrease and the normal operation of the system will be disturbed.

To prevent the possibility of this malfunctioning, additional control devices are introduced into the circuit. One of these devices is shown in Figure 6.14. A time relay R is connected in series with the contacts of switch S in the control circuit of coils C_1 and C_2 of the selectors (Figure 6.13). This relay has a normally closed contact, which breaks τ seconds after switch S closes, and almost instantaneously makes when the switch opens.

The time lag (τ) of the relay is so chosen that the relay is too slow to be energized during normal operation (hunting) of the system. If, however, a malfunction has caused q to follow for too long a time a unidirectional trend (increasing or decreasing), the relay is energized, is immediately de-energized, and then once again energized. The selectors consequently step

over the successive contacts and q is reversed. Normal operation of the system is thus restored.

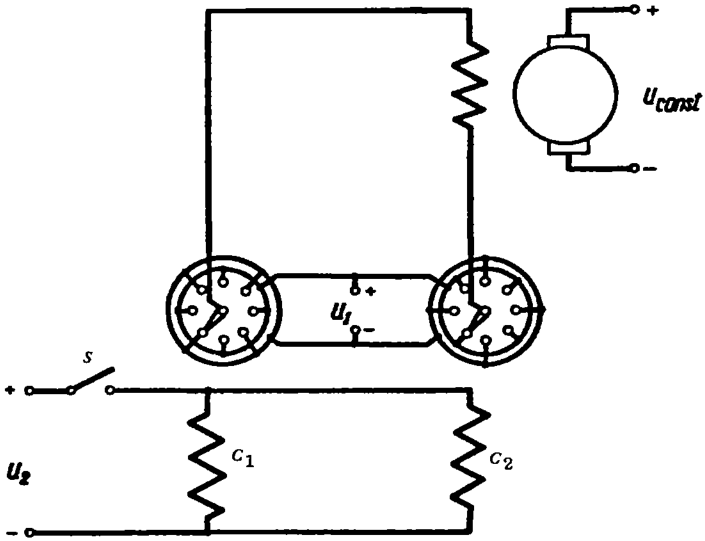

FIGURE 6.13. Circuit for reversal of fuel-feed motor.

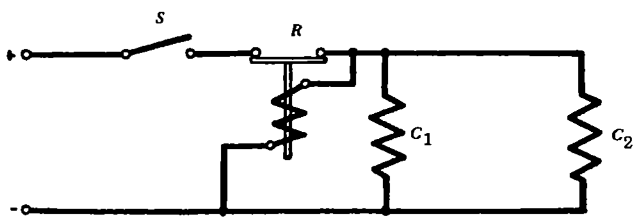

FIGURE 6.14. Self-control circuit.

6. EXTREMUM MEMORY SYSTEMS

The block diagram of an extremum memory system is shown in Figure 6.15,a. This system consists of a computing unit, for computing the extremal parameter (η), a comparator, a memory unit, and a control element (the analog of the fuel-fed motor in the previous example).

The basic circuit of the memory unit used in this system is shown in Figure 6.15,b. A signal u_1 proportional to the current value of the extremal parameter (η) is applied to the input of this unit. The output from this unit (Figure 6.15,a) is coupled to one of the inputs of the comparator. The output

of the comparator (u_3) is applied to the transistor control circuit. When the value of the extremal parameter (η) increases, the voltage at the output of the memory unit is proportional to this parameter $(u_2=u_1=k\eta)$.

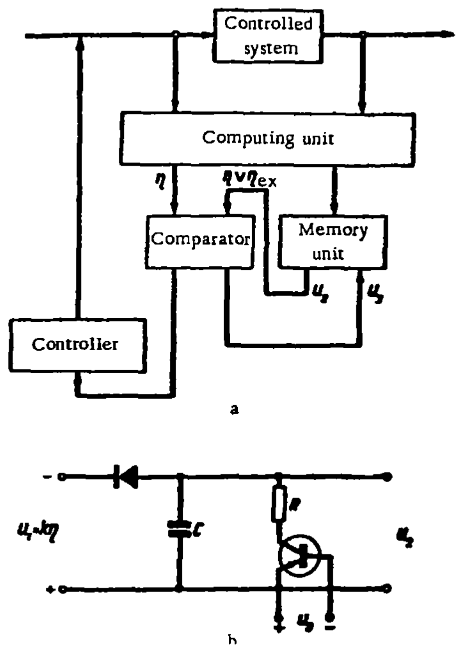

FIGURE 6.15. Extremum memory system.

When u_1 increases, the condenser C is charged. Let this charging be much more rapid than the rate of change of u_1. Then, as u_1 attains its maximum (extremum) value and starts decreasing, constant voltage equal to $u_{2\mathrm{ex}}$ is maintained at the input of the memory unit. The unit thus stores the extremum (maximum) value of the efficiency $(u_{2\mathrm{ex}}=k\eta_{\mathrm{ex}})$.

When a voltage u_3 is applied to the circuit, the condenser discharges through resistor R and the voltage at the circuit output is again equal to the input voltage $(u_2=u_1=k\eta)$.

The extremum memory system (Figure 6.15,a) functions as follows. During the increase of the extremal parameter, the output of the memory unit is proportional to the current of η. Both inputs of the comparator therefore receive equal voltage, both proportional to η, and no output signal is produced by this unit.

However, when the extremum is attained, the current value of the extremal variable (η) starts decreasing, whereas its maximum value of (η_{ex}) is stored by the memory unit and is maintained constant. Therefore, when the difference between these values of the extremal variable reaches a certain limit

$$\Delta\eta=\eta_{\mathrm{ex}_t}-\eta,$$

the comparator emits a signal which causes the correcting element to reverse the input signal (q).

On reversal, signal u_3 clears the memory (which has previously stored the maximum value η_{ex_i} of the extremal variable) and the output voltage of the storage is again proportional to the current value η. The process is resumed until the parameter η reaches its maximum value. At this point the memory unit stores a new value $\eta_{ex_{i+1}}$ and when the difference

$$\Delta\eta = \eta_{ex_{i+1}} - \eta$$

is attained, the system is again reversed. The system thus continually hunts the extremum.

Sign-sensing systems and extremum memory systems are relatively simple. These systems, however, are too sensitive to noise. The systems considered in the next section are free from this disadvantage.

7. STEP SYSTEMS

In these systems, the extremal variable is measured at definite time intervals, called *steps*. The values measured at the beginning and end of each interval (step) are then compared. If the extremal variable has decreased during the given step, the input parameter (q) is reversed. Otherwise, the system operates as normal.

The width of the step (Δt) is limited by the consideration that if too large a step is assumed, errors may creep in due to the drift of the extremal variable. On the other hand, if the step is too small, the normal functioning of the system can be affected by high-frequency noise generated in the system or received from outside.

Let us consider the step system in application to the previous example of a diesel generator, where the extremal variable is the efficiency . A block diagram of this system is shown in Figure 6.16.

In this system the extremal parameter η is determined, as before, by a computing unit. Since η continuously varies over the step, an integrator for producing the average value of the step (η_{av}) is coupled to the output of the computor. The integrator, the memory unit, and the comparator, all receive timing signals from a timer at intervals Δt.

The program followed by these devices when the time signal is received is shown in Table 6.1 (the 1st time signal is received Δt seconds after the diesel generator is started). In the first step, η_{av_1} is compared with zero (since the system has just started its operation). In subsequent steps the current η_{av_i} is compared with the previous value $\eta_{av_{i-1}}$. If

$$\eta_{av_i} > \eta_{av_{i-1}},$$

the system operates normally. Otherwise, the controlling element is actuated.

From Table 6.1 it can be seen that the memory unit consists of three stages operating in sequence, in the cyclic order 1-2-3-1, etc.

Let us now consider some characteristics of the step system. To simplify the discussion, we shall assume the rate of change of fuel supply $\left(\frac{dq}{dt}\right)$ to remain constant in time.

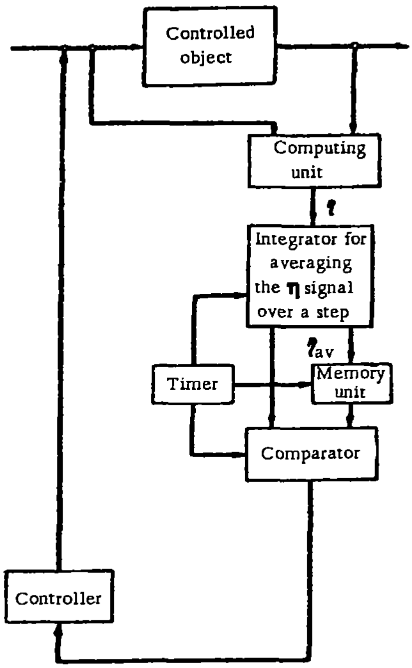

FIGURE 6.16. Step control system.

We should first note the following circumstances in Figure 6.17. The integrator computes the average value η_{av_i} during the i-th step, but produces the result only during the $(i+1)$-th step. The comparator therefore does not determine the current variation of η, but rather its variation during the preceding step. The position of the system relative to the extremum is thus established with a delay of one-step. The integrator, moreover, substitutes a discrete dependence of η on q for the continuous curve $\eta = f(q)$.

We plot this discrete (step) dependence (Figure 6.18) adjoining the graph $q = \varphi(t)$. We further mark on the step curve the dead zone of the comparator. This zone exists as there is always a certain finite difference $\eta_{av_{i-1}} - \eta_{av_i}$ to which the device should react.

Let us now follow the time variation of the quantity of fuel. First it increases (we have previously assumed the rate of change of fuel to be constant), but at point A the difference $\eta_{av_{i-1}} - \eta_{av_i}$ exceeds the dead zone limit of the comparator, the controller is actuated*, and the fuel supply decreases. At point B the system is again reversed, and the fuel supply increases. We thus see that as the system hunts near the extremum, the quantity of fuel fluctuates between q_{max} and q_{min}.

* It should be remembered that for the sake of simplicity here, as before, we neglect the time lag introduced by the controlled member.

TABLE 6.1

Step number	Output of averaging device	Storage device						Comparison device (compares)
		1st stage		2nd stage		3rd stage		
		stores	clears	stores	clears	stores	clears	
1	η_{av_1}	η_{av_1}	—	—	—	—	—	η_{av_1} c 0
2	η_{av_2}	—	—	η_{av_2}	—	—	—	η_{av_2} c η_{av_1}
3	η_{av_3}	—	η_{av_1}	—	—	η_{av_3}	—	η_{av_3} c η_{av_2}
4	η_{av_4}	η_{av_4}	—	—	η_{av_2}	—	—	η_{av_4} c η_{av_3}
5	η_{av_5}	—	—	η_{av_5}	—	—	η_{av_3}	η_{av_5} c η_{av_4}

Given the functions $q=\varphi(t)$ and $\eta=f(q)$, we may plot the curve of efficiency vs. time. This curve is shown in the bottom graph of Figure 6.18. From this graph it can be seen that hunting causes the efficiency to fluctuate between a maximum and a minimum value. The narrower the dead zone of the comparator and the narrower the timing signal, the smaller the fluctuations of η.

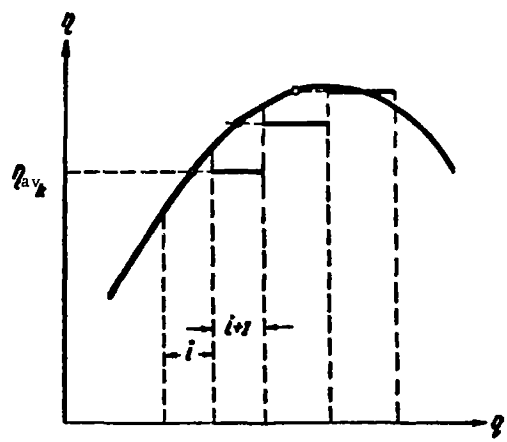

FIGURE 6.17. Operation characteristic of the averaging devices.

The step control system is highly reliable in operation, and in spite of its complexity, is very widespread.

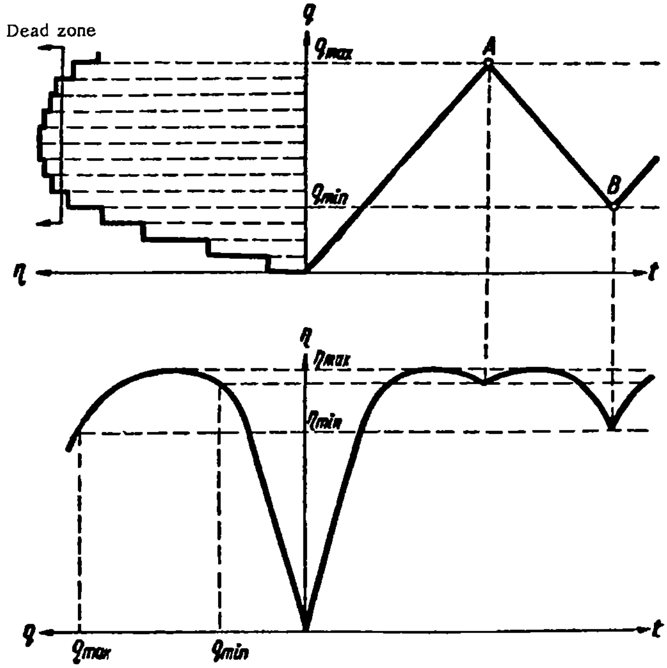

FIGURE 6.18. Characteristic curves of the step control system.

In step control systems, as in other types of adaptive control systems, the hunting of the extremum is regulated by analyzing the effect of the basic input signal (the quantity of fuel) on the result of operation of the system.

In addition, the position of the system relative to the extremum can be determined by the use of an auxiliary hunting signal superimposed on the basic signal at the system input. In this case the hunting of the extremum is regulated by analyzing the auxiliary signal produced at the system output when the hunting signal is fed to its input.

8. SYSTEMS WITH AN AUXILIARY HUNTING SIGNAL

A sinusoidal signal

$$u_1 = u_m \sin \omega t$$

is fed into the system.

It can be seen from Figure 6.19 that, when the system operates in the left part of the extremal characteristic (say, at point 1), the input (u_1) and the output (u_2) signals are out of phase by 180° (the positive halfwave of u_1 corresponds to the negative halfwave of u_2). If the system operates in the

right part of the extremal characteristic (say, at point 3), u_1 and u_2 are in phase.

At the extremum (point 2) the output signal does not contain the harmonic whose frequency is equal to that of the input signal (ω). Therefore, if a filter passing only the frequency of the input signal is provided at the output, a comparison of the phase of the transmitted signal with the phase of the input signal can be used to establish the branch of the extremal characteristic the system is currently operating, and in which direction the input variable (here, q) should be changed to attain the extremum.

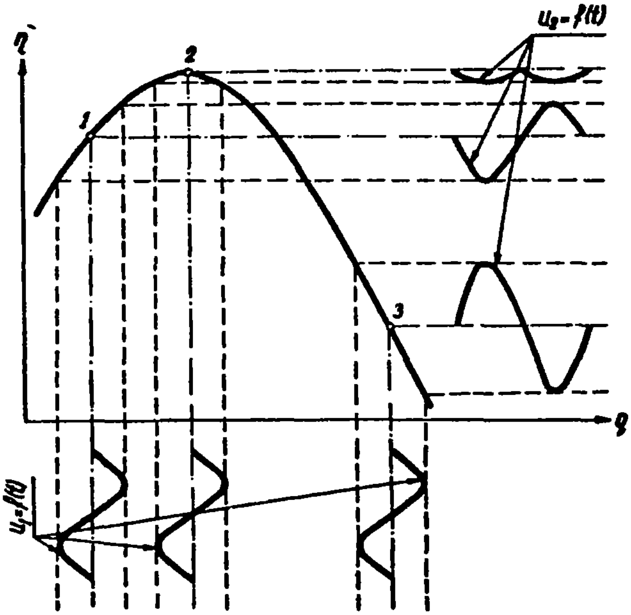

FIGURE 6.19. Transmission of signal through extremum system.

A block diagram of the system with an auxiliary continuous hunting signal is shown in Figure 6.20. The controlled system is, as previously, a diesel generator whose extremal characteristic is shown in Figure 6.19.

The computor in this system computes the efficiency of the diesel generator and the filter passes only the sinusoidal component of this coefficient whose frequency is equal to that of the hunting signal (ω).

A sine wave generator of frequency ω is connected to the input of the system. If the system is not tuned to the extremum, the filter output will be a sinusoidal voltage $\Delta\eta$. This voltage, together with the voltage Δu_1 produced by the generator, is applied to the inputs of an AND gate.

The circuit of the AND gate is shown in Figure 6.21. The signals Δu_1 and $\Delta\eta$ are applied to the inputs, so that when the system operates in the right part of the extremal characteristic, the polarity of these signals (during one half cycle) is as shown in the figure. The output of the AND gate (Δu_1 is much greater than $\Delta\eta$) is then a half wave voltage u_{AND} which can be used

as the reversal signal. When the system moves to the left part of the extremal characteristic there is a phase shift of 180° in voltage $\Delta\eta$, and the output voltage u_{AND} drops to zero. The output of the AND gate u_{AND} is delivered to the changeover switch (Figure 6.22).

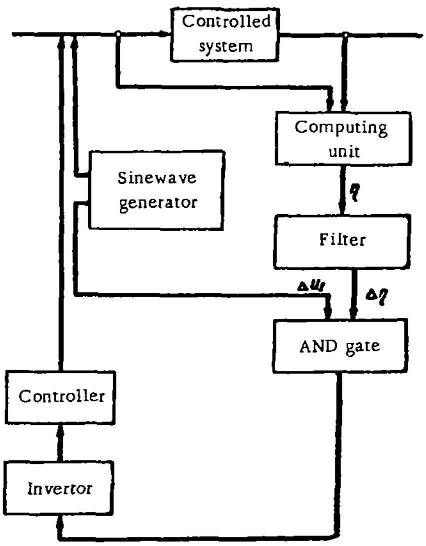

FIGURE 6.20. Block diagram of a control system with an auxiliary hunting signal.

Let us now consider the operation of the system in Figure 6.20. We switch in the diesel generator. The system operates in the left part of the extremal characteristic (Figure 6.19). The fuel-feed motor (controller) steps up the fuel supply (q) and the system moves toward the extremum. As soon as the system passes through this point, $\Delta\eta$ acquires a phase shift of 180° and the AND gate produces a reversal (u_{AND}). The signal operates the switch (Figure 6.22) and the polarity of the current in the excitation winding is reversed. The controller therefore reduces the quantity of fuel. When the system again passes through the extremum, the phase of $\Delta\eta$ changes and u_{AND} drops to zero. The polarity of the motor winding is again reversed and the fuel supply is stepped up.

Until now, in our discussion of the method with a continuous hunting signal, we neglected any possible time lag in the controlled system which could shift the phase of the output signal relative to the phase of the input signal. If there is time lag, the control system must contain a device compensating for phase shift. In the diagram of Figure 6.20 this device should be connected between the filter and the AND gate.

The phase shift arising when the signal is transmitted through the controlled system is variable because of the inconstancy of the parameters of the controlled system. Therefore, this compensation involves particular

difficulties, and consequently limits the applicability of continuous-hunting systems. This system is generally employed when the extremal characteristic drifts in jumps, as jumps of the extremal variable (η) do not produce false reversal signals (provided the system remains in the same portion of the extremal characteristic). In step systems, on the other hand, a discrete increment of the extremal characteristic (e.g., a small jumplike shift) can result in malfunctioning of the system.

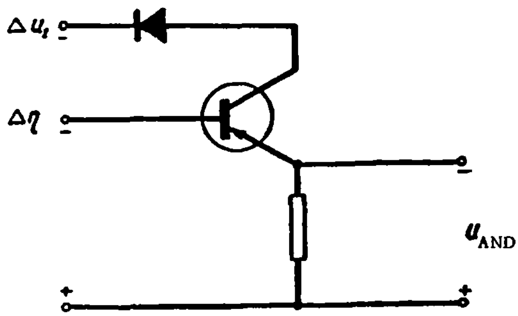

FIGURE 6.21. AND gate circuit.

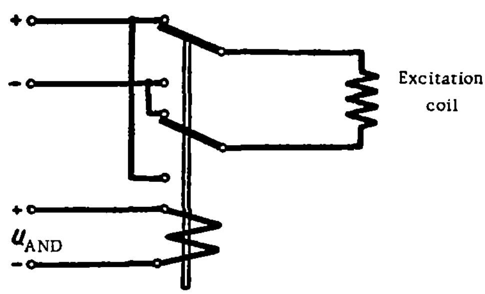

FIGURE 6.22. Circuit for polarity reversal of the control motor.

9. PRINCIPAL PARAMETERS CHARACTERIZING EXTREMUM CONTROL

The extremal characteristic is translated so that the extremum coincides with the origin (Figure 6.23). This simplifies its mathematical description. The characteristic shown in the figure has a maximum. All that follows, however, applies equally well to a characteristic with a minimum.

Although our reasoning applies to any extremum system, we shall retain the notation η for the extremal variable, and q for the input variable which determines the functional variation of η.

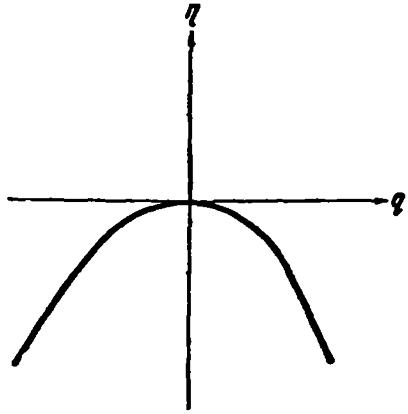

FIGURE 6. 23. Extremal characteristic.

We shall assume, as before, that the rate of change of the input variable q is constant. Moreover, we shall assume that near the extremum the characteristic is parabolic, and is described by the equation

$$\eta = -k_1 q^2, \tag{6.3}$$

where k_1 is a coefficient of proportion.

The control process starts from the time when q is zero*. First (Figure 6.24) A increases, but at point q the system is reversed and q starts decreasing. At point B another reversal takes place and q again increases. The parameter q thus hunts (oscillates periodically) about the zero value (the extremum point).

From (6.3), it can be seen that the variable η also hunts, but its period is equal to one half of the period of q (Figure 6.24). This period (T) is called the hunting period. The maximum deviation of η from the extremum is called the hunting amplitude of the output and is denoted as Δ (Figure 6.24).

In Figure 6.24 it can be seen that during the interval $\frac{1}{2}T > t > 0$ the input variable q varies as

$$q = kt; \left(\frac{1}{2}T > t > 0\right), \tag{6.4}$$

where $k = \operatorname{tg}\alpha$ is the slope of the characteristic $q = \varphi(t)$ relative to the horizontal axis.

Inserting this q into (6.3), we obtain

$$\eta = -k_1 k^2 t^2; \left(\frac{1}{2}T > t > 0\right) \tag{6.5}$$

* It should be observed that since the extremum (Figure 6.23) has been transferred to the origin, q can be either positive or negative.

The rate of change of the extremal variable in the interval $\frac{1}{2}T > t > 0$ is therefore given by

$$\frac{d\eta}{dt} = -2k_1k^2t. \qquad (6.6)$$

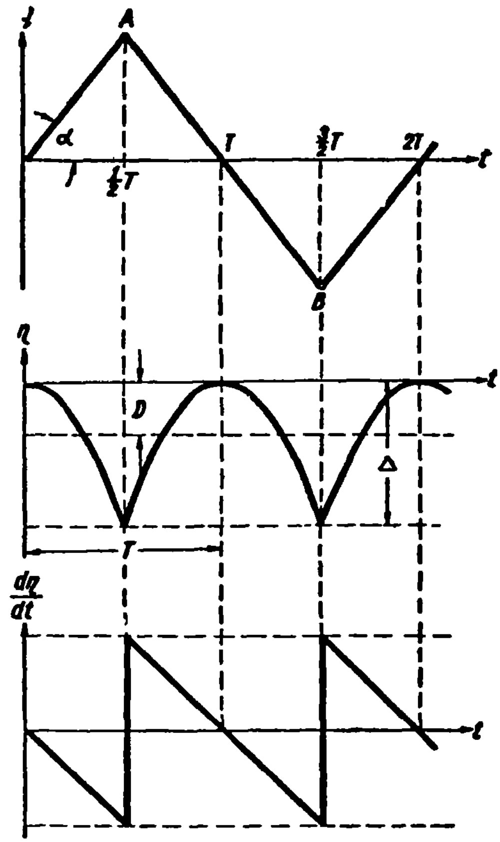

FIGURE 6.24. Characteristics of extremum control.

This function is shown in the bottom graph of Figure 6.24. At $t = \frac{1}{2}T$ the system is reversed and the derivative of the extremal variable changes its sign. Further variation of the derivative is evident from Figure 6.24. The maximum value of this derivative where the system is reversed is called the critical value.

The average value D of the extremal variable over one hunting period is called the hunting loss and is defined as

$$D=\frac{1}{T}\int_{0}^{T}\eta dt,$$

where T is the hunting period.

Inserting η from (6.5) and changing the limits of integration, we obtain

$$D=-\frac{2k_1k^2}{T}\int_{0}^{\frac{1}{2}T}t^2dt=-\frac{k_1k^2T^2}{12} \tag{6.7}$$

It should be observed that since the hunting loss is an integral quantity, it is equal to the d-c component of the extremal variable (η), as the average of all the harmonic components over one period is obviously zero.

The hunting amplitude (Figure 6.24) is derived from (6.5) when we set $t=\frac{1}{2}T$:

$$\Delta=-\frac{k_1k^2T^2}{4}. \tag{6.8}$$

Comparing equations (6.7) and (6.8), we see that the hunting amplitude is equal to three times the hunting loss:

$$\Delta=3D. \tag{6.9}$$

It follows from Figure 6.24 that the critical value of the derivative of the extremal variable (6.6) is equal to $(t=\frac{1}{2}T)$:

$$\left(\frac{d\eta}{dt}\right)_{max}=-k_1k^2T. \tag{6.10}$$

Solving equations (6.7) and (6.10) simultaneously, we obtain

$$D=-\frac{1}{12k_1k^2}\left(\frac{d\eta}{dt}\right)^2_{max}.$$

Hence, to reduce the hunting losses the critical value of the derivative must be lowered. This reduction, however, is limited when false reversals of the system are considered.

The hunting period must also be made as small as possible, since this reduces the time taken by the system to arrive at the extremum. Too small a hunting period, however, makes it difficult to distinguish between the variation of the input parameter (q) due to hunting, and increments produced by random processes (e.g., high-frequency noise).

10. THE EFFECT OF TIME LAG ON THE CHARACTERISTICS OF EXTREMUM CONTROL

Until now, in our discussion of the extremal characteristics of adaptive control systems, we assumed all the component elements to be intertia-less. Since real systems have inertia, this effect must be allowed for.

To simplify the analysis we divide the adaptive extremum system into several parts (Figure 6.25). In Figure 6.25 the system is divided into three parts: a linear input unit with transfer function $G_1(s)$, a nonlinear unit, and a linear output unit with transfer function $G_2(s)$. The division is made so that all the intertial elements of the system are lumped in the input and output units, while the nonlinear unit remains inertia-less.

The input unit generally actuates the control element and the input amplifiers. The output unit, as a rule, comprises the output device, the smoothing filter, etc. It is also assumed that the extremum hunting unit, used here as the feedback loop, is nonlinear, albeit inertia-less. The object of this hunting unit is to set up an alternating signal of constant amplitude (u_1):

$$\left.\begin{array}{l} u_1=+k \text{ (when } q \text{ is to be increased):} \\ u_1=-k \text{ (when } q \text{ is to be lowered):} \end{array}\right\} \tag{6.11}$$

A constraint imposed on the transfer functions of the linear elements, $G_1(s)$ and $G_2(s)$, is that the time lag introduced by these functions must be small in comparison with the hunting period of the system in the inertia-less case. Otherwise the normal operation of the system is disturbed.

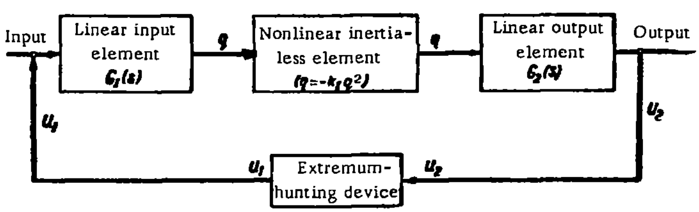

FIGURE 6.25. Block diagram of an adaptive extremum system.

Since most real input and output units of such systems can be described by linear differential equations, we shall confine the following discussion to this class only.

Let us first consider a particular case when the lag is produced at the output only. In other words, we take $G_1(s)=1$. The time dependence of the input (q) and the extremal (η) variables is shown in Figure 6.24. Since in our case all the functions remain the same, we simply reproduce the previous graph in Figure 6.26. Here, however, η is not the output variable of the system. When passing through the output element the signal η is smoothed and acquires a phase shift. The output signal of the system therefore has the form $u_2=\varphi(t)$, shown graphically in Figure 6.26.

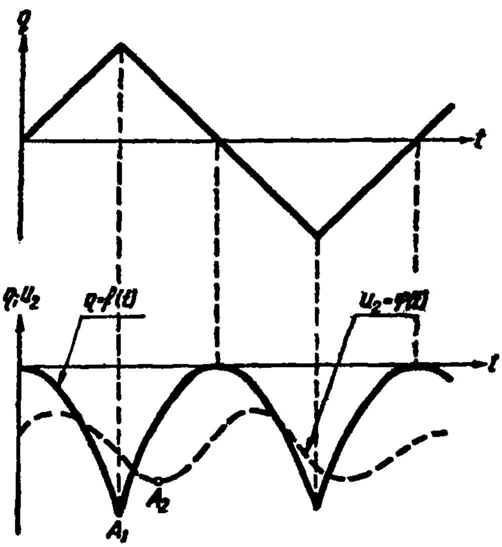

FIGURE 6. 26. Characteristics of an extremum system with the time lag restricted to the output.

Hence, a time lag at the system output causes the minimum of the input parameter (u_2) of the system (point A_2) to lag behind the true time of transmission of the minimum value of the extremal variable (point A_1). This introduces an error in the determination of the reversal point. To eliminate this error, a phase advance element must be provided at the system output.

We shall now consider a more general case, when time lags arise both at the input and at the output.

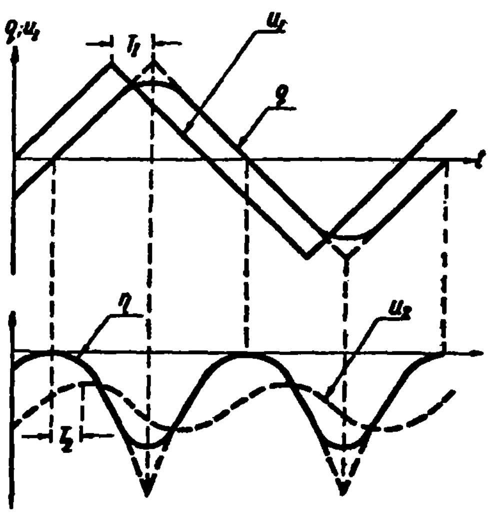

FIGURE 6. 27. Characteristics of an extremum system with time lags arising at the input and the output.

In this case the input signal u_1 (Figure 6.27) is affected by the reversal signal like q in the intertia-less case (Figure 6.24). The signal u_1 acquires a phase shift upon passing through the input unit and its curve is somewhat smoothened $[q=F(t)]$. The critical value of the extremal variable $[\eta=f(t)]$ is therefore somewhat reduced and varies as shown in Figure 6.27. The output function u_2, as in the case where the time lag was restricted to the output (Figure 6.26), shifts relative to the $\eta=f(t)$ curve and is smoother than this extremal curve.

Calculations show that the time lag introduced by the input (T_1) and the output (T_2) units of the system is in each case approximately equal to the time constant of the respective units.

11. THE DYNAMICS OF EXTREMUM HUNTING IN ADAPTIVE CONTROL SYSTEMS

Let us consider the dynamics of extremum hunting in an adaptive control system as it starts from the zero point and tunes itself to the extremum of the controlled variable.

Suppose that the input linear unit (Figure 6.25) is an integrator, i.e., signal q steadily increases when signal u_1 is positive (see (6.11)), and decreases when u_1 is negative. For the input unit we have, therefore, from (6.4)

$$\frac{dq}{dt}=u_1=\pm k, \tag{6.12}$$

where k is a constant.

The output linear unit of the system is described, as we have already observed, by a linear differential equation. Its transfer function therefore has the form

$$G_2(s)=\frac{k_2}{1+sT_2}, \tag{6.13}$$

where T_2 = the time constant of the output unit;

k_2 = the gain of this unit.

Hence, the ratio between the Laplace transforms of the signals at the output and the input of the output unit is equal to

$$U_2(s)=\frac{k_2}{1+sT_2}H(s)^*,$$

where $H(s)$ is the Laplace transform of $\eta(t)$.

* Here, as before, the Laplace transforms $U_2(s)$ and $H(s)$ are denoted by capital letters, to distinguish from the inverse functions $u_2(t)$ and $\eta(t)$.

We rewrite this equation in a somewhat different form:

$$sU_2(s) = \frac{1}{T_2}[k_2H(s) - U_2(s)].$$

The inverse function of this equation is

$$\frac{du_2}{dt} = \frac{1}{T_2}(k_2\eta - u_2). \tag{6.14}$$

We further assume that the system is reversed by the extremum hunting unit when the error of the system output signal u_2 equals

$$\Delta u_2 = u_{2max} - u_2, \tag{6.15}$$

where u_{2max} = the signal extremum at system output;
Δu_2 = the prescribed error of output voltage relative to the extremum.

Solving equations (6.12) and (6.14) simultaneously, we obtain

$$\frac{du_2}{dq} = \pm\frac{1}{kT_2}(k_2\eta - u_2). \tag{6.16}$$

It follows from this equation that the maximum signal (u_2) is attained at the output when

$$u_2 = k_2\eta; \left(\frac{du_2}{dq} = 0\right). \tag{6.17}$$

If now the extremal parameter is appropriately scaled (setting $\eta_s = k_2\eta$), the signal u_2 is maximized when

$$u_2 = \eta_s, \tag{6.17a}$$

i.e., at the intersection points of the curves $u_2 = \varphi(q)$ and $\eta_s = f(q)$.

Proceeding from these considerations, we shall now analyze the dynamics of extremum hunting in this system.

Let the system be at the origin (Figure 6.28, point a). When q is increasing, the following inequality applies near point a owing to the time lag in the output unit:

$$\eta_s > u_2, \tag{6.18}$$

i.e., curve $u_2 = \varphi(q)$ extends below curve $\eta_s = f(q)$.

Since q increases, we have from equations (6.16) and (6.18) in the vicinity of point a

$$\left(\frac{du_2}{dq}\right)_a > 0.$$

This inequality will persist till point b is reached, where the curves $u_2 = \varphi(q)$

and $\eta_s = f(q)$ intersect. Therefore, from (6.17a), at point b

$$\left(\frac{du_2}{dq}\right)_b = 0.$$

Passing through the maximum at point b, u_2 starts decreasing. As soon as the difference $u_{2_b} - u_2$ becomes equal to Δu_2 (see (6. 15)), the system is reversed (point c). The input signal u_1 changes its sign and from (6.11)

$$(u_1)_{c-f} = -k$$

and q starts decreasing.

Beyond point c the $u_2 = \varphi(q)$ curve first extends above the $\eta_s = f(q)$ curve. Therefore $(u_2 > \eta_s)$, seeing that the right-hand side in (6.16) is negative $(u_1 = -k)$, we obtain

$$\left(\frac{du_2}{dq}\right)_{c-d} > 0,$$

i.e., the slope of the $u_2 = \varphi(q)$ curve is positive.

Since beyond point c the variable q decreases and the function $u_2 = f(q)$ has a positive slope, u_2 also decreases. At a certain point (d), the curves $u_2 = \varphi(q)$ and $\eta_s = f(q)$ will therefore meet. Since at point d, $u_2 = \eta_s = k_2\eta$, the derivative (6.16) is equal at this point to

$$\left(\frac{du_2}{dq}\right)_d = 0.$$

Beyond point d, the signal u_2 is lower than η_s. Therefore, according to (6.16), the curve $u_2 = \varphi(q)$ has again a negative slope $(u_1 = -k)$:

$$\left(\frac{du_2}{dq}\right)_{d-e} < 0.$$

At point e this derivative vanishes again. Beyond point e the voltage u_2 again decreases, since q is still decreasing and the derivative $(u_2 > \eta_s)$ is positive:

$$\left(\frac{du_2}{dq}\right)_{e-f} > 0.$$

At point f the difference $u_{2_e} - u_2$ becomes equal to Δu_2 and the system is again reversed, so that q increases. This pattern will persist until the system has approached the extremum and starts hunting about the extremum point.

The curve $u_2 = \varphi(q)$ describes the steady hunting mode called the l i m i t h u n t i n g c y c l e (Figure 6.29). A characteristic feature of the limit cycle distinguishing it from the dynamic extremum-hunting cycle (Figure 6.28) is that any point of the cycle (e.g., point A in Figure 6.29) repeats itself when two hunting periods have been completed (for point A this indicates motion from A to B and back).

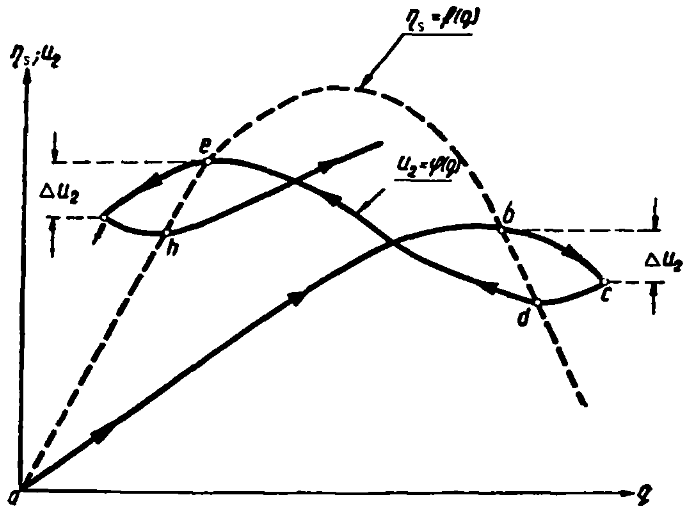

FIGURE 6.28. The dynamics of extremum hunting.

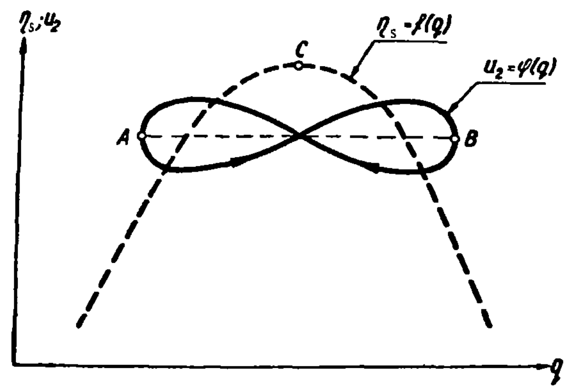

FIGURE 6.29. Limit hunting cycle.

The system operates in the limit cycle as long as the extremum C does *not drift. When this point drifts, the system tunes intself to the new ex*tremum and a new limit cycle is established.

Let us now consider the mathematics of the dynamic extremum hunting (Figure 6.28) and of the limit cycle (Figure 6.29). To simplify the discussion we shall assume a parabolic extremal characteristic $\eta = f(q)$. A method similar to the following can be used to trace the extremum hunting of a system with any other extremal characteristic.

From Figure 6.28 it can be seen that the dynamic hunting curve of an adaptive system $u_2 = \varphi(q)$ is in fact a succession of hunting halfcycles. Over a positive halfcycle, q increases. The positive halfcycle is then succeeded by a negative halfcycle and q decreases. Completing several alternating halfcycles the system tunes to the extremum and settles in a limit hunting

cycle (Figure 6.29). To describe the behavior of the system, it is sufficient to determine the equations of the system for a positive and a negative half-cycle.

Let us first consider a positive halfcycle originating at an arbitrary point (u_2, q_0).

Since over a positive halfcycle (see (6.11)) the input signal u_1 (Figure 6.25) is $+k$, then, inserting η from (6.3) into (6.16), we obtain

$$\frac{du_2}{dq}=-\frac{1}{kT_2}(k_1k_2q^2+u_2)$$

or

$$\frac{du_2}{dq}+\frac{1}{kT_2}u_2=-\frac{k_1k_2}{kT_2}q^2. \tag{6.19}$$

Solving this equation, we derive the relationship required.

We have assumed nonzero initial conditions (u_{2_0} and q_0) in this equation (halfcycle). Therefore, to reduce the system to one with zero initial conditions, we substitute the variables

$$\left.\begin{array}{l} z=u_2-u_{2_0}; \\ \tau=q-q_0. \end{array}\right\} \tag{6.20}$$

Then $dz=du_2$, $d\tau=dq$, and therefore

$$\frac{du_2}{dq}=\frac{dz}{d\tau}. \tag{6.20a}$$

Inserting u_2, q from (6.20) and $\frac{du_2}{dq}$ from (6.20a) into (6.19), we obtain

$$\frac{dz}{d\tau}+\frac{1}{kT_2}z=-\frac{1}{kT_2}(k_1k_2\tau^2+2k_1k_2q_0\tau+u_{2_0}+k_1k_2q_0^2). \tag{6.21}$$

This expression differs from (6.19) in that it has zero initial conditions. Indeed, it follows from (6.20) that when $q=q_0$, the variable $\tau=0$. On the other hand, for $q=q_0$, the signal $u_2=u_{2_0}$. We therefore have from (6.20), $z=0$.

Since (6.21) has zero initial conditions, we Laplace-transform the variables z and τ and write

$$\left(s+\frac{1}{kT_2}\right)Z(s)=-\frac{1}{kT_2}\left(\frac{2k_1k_2}{s^3}+\frac{2k_1k_2q_0}{s^2}+\frac{u_{2_0}+k_1k_2q_0^2}{s}\right)$$

or

$$Z(s)=-\frac{1}{kT_2}\cdot\frac{(u_{2_0}+k_1k_2q_0^2)s^2+2k_1k_2q_0s+2k_1k_2}{s^3\left(s+\frac{1}{kT_2}\right)}.$$

Rewriting the right-hand side of this equation as a sum of partial fractions, we obtain

$$Z(s)=-\frac{2k_1k_2}{s^3}-\frac{2k_1k_2(q_0-kT_2)}{s^2}-\frac{u_{2_0}+k_1k_2(q_0^2-2q_0kT_2+2k^2T_2^2)}{s}+$$
$$+\frac{u_{2_0}+k_1k_2(q_0^2-2q_0kT_2+2k^2T_2^2)}{s+\frac{1}{kT_2}}.$$

Using the Laplace-transform tables, we find the inverse function

$$z=-k_1k_2\tau^2-2k_1k_2(q_0-kT_2)\tau-u_{2_0}-k_1k_2(q_0^2-2q_0kT_2+2k^2T_2^2)+$$
$$+\left[u_{2_0}+k_1k_2(q_0^2-2q_0kT_2+2k^2T_2^2)\right]e^{-\frac{\tau}{kT_2}}.$$

Inserting z and τ from (6.20) and carrying out some simple manipulations gives

$$u_2=-k_1k_2(q^2-2kT_2q+2k^2T_2^2)+$$
$$+\left[u_{2_0}+k_1k_2(q_0^2-2q_0kT_2+2k^2T_2^2)\right]e^{-\frac{q-q_0}{kT_2}}. \tag{6.22}$$

This is the equation of a positive halfcycle of the system, when the parameter q increases. To obtain the equation of a negative halfcycle, it suffices to change the sign of the coefficient k in (6.19), and consequently in (6.22), since according to (6.11), $u_1=-k$ over the negative halfcycle.

Equation (6.22) for a negative halfcycle can therefore be written as

$$u_2=-k_1k_2(q^2+2kT_2q+2k^2T_2^2)+$$
$$+\left[u_{2_0}+k_1k_2(q_0^2+2q_0kT_2+2k^2T_2^2)\right]e^{\frac{q-q_0}{kT_2}}. \tag{6.23}$$

The entire curve characterizing the dynamics of extremum tuning is plotted as follows.

The origin is translated to the extremum (Figure 6.23). When the system is started, the parameters u_{2_0} and q_0 are therefore not zero, as in Figure 6.28, but rather u_{2_a} and q_a. (Figure 6.30). Inserting $u_{2_0}=u_{2_a}$ and $q_0=q_a$ in (6.22), we obtain the equation for the first positive halfcycle of system motion:

$$u_2=-k_1k_2(q^2-2kT_2q+2k^2T_2^2)+$$
$$+\left[u_{2_a}+k_1k_2(q_a^2-2q_akT_2+2k^2T_2^2)\right]e^{-\frac{q-q_a}{kT_2}}. \tag{6.24}$$

We plot this curve in Figure 6.30. The signal first increases, but at

point b it is arrested and starts to decrease. When u_2 equals u_{2c}, defined as

$$u_{2_c} = u_{2_b} - \Delta u_2, \tag{6.25}$$

the system is reversed. Point c is therefore the origin of the first negative halfcycle of hunting.

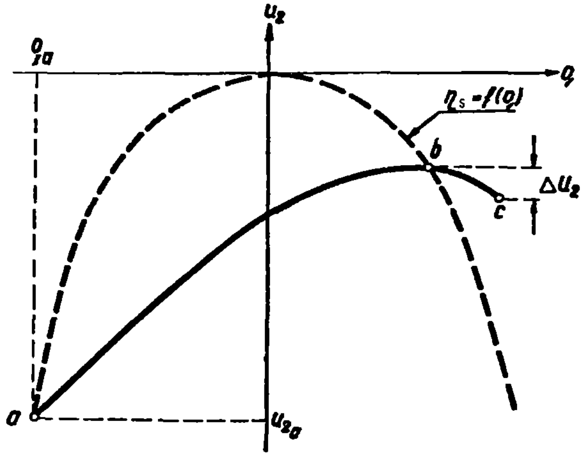

FIGURE 6.30. Characteristic of the system with the origin translated to the extremum.

The coordinates of point c are determined as follows. Point b is the intersection point of curves $\eta_s = f(q)$ and $u_2 = \varphi(q)$ (here $u_{2_b} = \eta_{s_b}$). Therefore, equating the right-hand side of (6.24) with the expression $\eta_s = -k_1k_2q^2$ (see (6.3) and footnote on p. 209) for $q = q_b$, we obtain the following equation:

$$2k_1k_2kT_2(kT_2 - q_b) = [u_{2_a} + k_1k_2(q_a^2 - 2q_a kT_2 + 2k^2T_2^2)]e^{-\frac{q_b - q_a}{kT_2}}. \tag{6.26}$$

This equation can be solved for q_b. Given q_b, we can find from (6.24) the second coordinate of point b, u_{2_b}. Now, since Δu_2 is known, u_{2_c} can be determined from (6.25). Given u_{2_c}, we can obtain q_c from (6.24). Inserting the initial values $q_0 = q_c$ and $u_{2_0} = u_{2_c}$ into (6.23), we obtain the equation of the first negative halfcycle.

Repeating this procedure, the entire dynamic curve describing the tuning of the system to the extremum, is plotted.

Example. An external characteristic of a system is

$$\eta = -q^2 \text{ (i. e., } k_1 = 1)$$

and the coordinates of its point of origin, $u_{2_a} = -9$ v, $q_a = -3$ v, are given.

Moreover, $k=1$, $k_2=1$, $T_2=2$ sec, $\Delta u_2=0.25$v.

Inserting these parameters in (6.24) gives the equation of the first positive halfcycle:

$$u_2=-q^2+4q-8+20e^{-0.5q-1.5}. \tag{6.27}$$

To determine the coordinate q_b, the parameters are inserted in (6.26), giving

$$q_b=2-5\,e^{-0.5q_b-1.5}.$$

Solving this equation (by trial and error) for q_b, we find

$$q_b=1.5\text{ v.}$$

Inserting q_b in (6.27), the voltage at point b is

$$u_{2_b}=-2.24\text{ v.}$$

We further determine the coordinates of point c. Inserting u_{2_b} and Δu_2 in equation (6.25) we find

$$u_{2_c}=u_{2_b}-\Delta u_2=-2.24-0.25=-2.49\text{ v.}$$

From equation (6.27), where $u_{2c}=-2.49$v, we obtain

$$q_c=2.14\text{v.}$$

Point c with the coordinates $q_c=2.14$v, $u_{2c}=-2.49$v is the point of origin of the first negative halfcycle. Inserting the various parameters into (6.23) and seeing that $q_0=q_c=2.14$v and $u_{2_0}=u_{2_c}=-2.49$v, we obtain the equation of the first negative halfcycle:

$$u_2=-q^2-4q-8+18.65\,e^{0.5q-1.07}.$$

We now can find the end point of this negative halfcycle. This point is obviously the point of origin of the second positive halfcycle, etc. Repeating this procedure, the entire dynamic curve of extremum tuning is plotted.

Equations (6.22) and (6.23) apply to any halfcycle of system motion, including the limit halfcycles.

A positive halfcycle coincides with the positive limit halfcycle if it passes through points A, B, and C (Figure 6.31). Note that points A and C have equal ordinates $(u_{2_A}=u_{2_C})$ and are equally removed from the ordinate axis $(q_A=-q_C)$. Point B is the intersection point of curves $\eta_s=f(q)$ and $u_2=\varphi(q)$. These conditions enable us to determine the equations of the limit halfcycles.

The point of origin of the positive limit halfcycle (Figure 6.31) is point A. Therefore, inserting the initial conditions $u_{2_0}=u_{2_A}$, $q_0=q_A$ in (6.22), the

characteristic of the positive limit halfcycle is

$$u_2 = -k_1k_2(q^2 - 2kT_2q + 2k^2T_2^2) +$$
$$+\left[u_{2_A} + k_1k_2(q_A^2 - 2q_A kT_2 + 2k^2T_2^2)\right] e^{-\frac{q-q_A}{kT_2}}. \qquad (6.28)$$

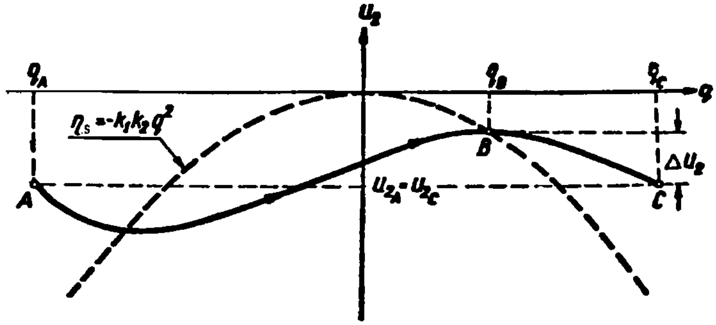

FIGURE 6. 31. Parameters of the limit cycle.

Two unknown parameters enter this equation, namely the coordinates of point A (u_{2_A} and q_A). We thus require additional conditions defining these parameters. We have previously observed that the curve of the positive limit halfcycle passes through point C. Therefore, inserting in (6.28) $q = q_C$, we obtain

$$u_{2_c} = -k_1k_2(q_c^2 - 2kT_2q_c + 2k^2T_2^2) +$$
$$+\left[u_{2_A} + k_1k_2(q_A^2 - 2q_AkT_2 + 2k^2T_2^2)\right] e^{-\frac{q_C - q_A}{kT_2}}. \qquad (6.29)$$

Point B (Figure 6.31) is the intersection point of the curves $\eta_s = f(q)$ and $u_2 = \varphi(q)$. For this point we therefore have from (6. 3) and (6. 17)

$$u_{2_B} = k_2\eta_B = -k_1k_2q_B^2. \qquad (6.30)$$

From (6.25)

$$u_{2C} = u_{2_B} - \Delta u_2,$$

therefore, applying (6.30), we obtain

$$u_{2C} = -k_1k_2q_B^2 - \Delta u_2. \qquad (6.31)$$

Moreover, inserting $q=q_B$ in (6.28), we write

$$u_{2_B} = -k_1k_2(q_B^2 - 2kT_2q_B + 2k^2T_2^2) + \\ + \left[u_{2_A} + k_1k_2(q_A^2 - 2q_A kT_2 + 2k^2T_2^2)\right]e^{-\frac{q_B - q_A}{kT_2}}.$$

Inserting u_{2_B} from (6.30) and q_B from (6.31) we obtain

$$2k_1k_2kT_2\left(kT_2 - \sqrt{\frac{-u_{2_C} - \Delta u_2}{k_1k_2}}\right) = \\ = \left[u_{2_A} + k_1k_2(q_A^2 - 2q_A kT_2 + 2k^2T_2^2)\right] e^{\frac{q_A - \sqrt{\frac{-u_{2_C} - \Delta u_2}{k_1k_2}}}{kT_2}}. \tag{6.32}$$

We have previously observed that a limit halfcycle differs from other halfcycles in that the coordinates of its two end points (A and C) are related by the equations

$$\left.\begin{array}{l} q_C = -q_A; \\ u_{2C} = u_{2A}. \end{array}\right\} \tag{6.33}$$

Therefore, inserting these conditions in (6.29) and (6.32), we derive

$$u_{2_C} = -k_1k_2(q_C^2 - 2kT_2q_C + 2k^2T_2^2) + \\ + \left[u_{2_C} + k_1k_2\,(q_C^2 + 2q_C kT_2 + 2k^2T_2^2)\right] e^{-\frac{2q_C}{kT_2}}; \tag{6.34}$$

$$2k_1k_2kT_2\left(kT_2 - \sqrt{\frac{-u_{2_C} - \Delta u_2}{k_1k_2}}\right) = \\ = \left[u_{2C} + k_1k_2(q_C^2 + 2q_C kT_2 + 2k^2T_2^2)\right]e^{-\frac{q_C + \sqrt{\frac{-u_{2_C} - \Delta u_2}{k_1k_2}}}{kT_2}}. \tag{6.35}$$

The solution of these simultaneous equations gives the coordinates of C(q_C and u_{2_C}). Now, from (6.33) we find the coordinates of point A. Applying (6.28), we can now plot the positive limit halfcycle. If in (6.28) we change the sign of k and substitute new initial conditions (q_C and u_{2C}) for q_A and u_{2A}, we derive the characteristic of the negative limit halfcycle:

$$u_2 = -k_1k_2(q^2 + 2kT_2q + 2k^2T_2^2) + \\ + \left[u_{2_C} + k_1k_2(q_C^2 + 2q_C kT_2 + 2k^2T_2^2)\right] e^{\frac{q - q_C}{kT_2}}. \tag{6.36}$$

Example. Determine the characteristic of the limit halfcycles for the parameters of the previous example: $k=1, k_1=1, k_2=1, T_2=2$ sec, $\Delta u_2=0.25$ v.

Inserting these parameters in equations (6.34) and (6.35), we obtain

$$\left.\begin{aligned} &u_{2C}=-q_C^2+4q_C-8+(u_{2_C}+q_C^2+4q_C+8)\,e^{-q_C};\\ &4\left(2-\sqrt{-u_{2_C}-0.25}\right)=\left(u_{2_C}+q_C^2+4q_C+8\right)e^{-0.5\left(q_C+\sqrt{-u_{2_C}-0.25}\right)}. \end{aligned}\right\} \tag{6.37}$$

Solving by trial and error, we find

$$q_C=1.47\,\text{v};\ u_{2_C}=-0.77\,\text{v}.$$

From (6.33) the coordinates of point A are

$$q_A=-1.47\,\text{v};\ u_{2_A}=-0.77\,\text{v}.$$

Inserting the parameters k, k_1, k_2, and T_2, and the initial conditions q_A and u_{2_A} into (6.28), the characteristic of the positive limit halfcycle is determined:

$$u_2=-q^2+4q-8+15.27e^{-0.5q-0.74}.$$

Similarly, inserting these data into (6.26), we find the characteristic of the negative limit halfcycle:

$$u_2=-q^2-4q-8+15.27e^{0.5q-0.74}.$$

Let us further consider the time variation of the output voltage (Figure 6.25) when the system operates in the limit hunting cycles. We shall assume, as before, that the rate of change of the variable q remains constant (in magnitude). Then, for the limit hunting cycle (Figure 6.31) q will follow the graph shown in Figure 6.32. According to this graph, the time dependence of q during the first hunting period (the positive halfcycle) is given by

$$q=q_A+kt;\ (T\geqslant t\geqslant 0), \tag{6.38}$$

where $k=\operatorname{tg}\alpha$.

Inserting this expression for q into (6.28) for the positive limit halfcycles we find

$$u_2=-k_1k_2[(q_A+kt)^2-2kT_2(q_A+kt)+2k^2T_2^2]+$$
$$+[u_{2_A}+k_1k_2(q_A^2-2q_AkT_2+2k^2T_2^2)]\,e^{-\frac{t}{T_2}}$$

or after simple manipulations

$$u_2=-kk_1k_2[kt^2+2(q_A-kT_2)t]+u_{2_A}e^{-\frac{t}{T_2}}+$$
$$+k_1k_2(q_A^2-2q_AkT_2+2k^2T_2^2)\left(e^{-\frac{t}{T_2}}-1\right). \tag{6.39}$$

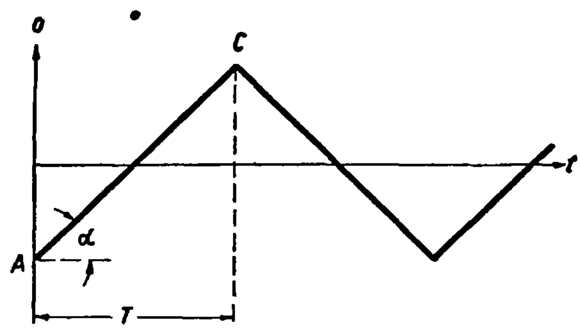

FIGURE 6.32. The characteristic of q.

The time variation of the extremal variable η can be derived by inserting q from (6.38) into (6.3). This gives

$$\eta = -k_1(q_A + kt)^2. \qquad (6.40)$$

It was shown in section 6-9 that the hunting loss D is defined as the average of the extremal variable over a hunting period:

$$D = \frac{1}{T}\int_0^T \eta\, dt. \qquad (6.41)$$

We first find the hunting period T. From Figures 6.31 and 6.32 it can be seen that the hunting period of a system is equal to the time required to move from point A to point C (or back). Therefore, setting $q = q_C$ in (6.38) we find

$$q_C = q_A + kT,$$

whence

$$T = \frac{1}{k}(q_C - q_A).$$

Inserting q_C from (6.33), we find the hunting period:

$$T = -\frac{2q_A}{k}. \qquad (6.42)$$

Inserting η from (6.40) and T from (6.42) into (6.41), we obtain

$$D = \frac{kk_1}{2q_A}\int_0^{-\frac{2q_A}{k}} (q_A + kt)^2\, dt,$$

whence, integrating, we find the hunting loss:

$$D=-\frac{1}{3}k_1q_A^2. \tag{6.43}$$

In the previous example, the hunting loss was

$$D=-\frac{1}{3}\cdot 1\cdot 1.47^2=0.72\text{ v}.$$

12. STABILITY OF ADAPTIVE CONTROL SYSTEMS

We have previously assumed a restriction to the effect that the response time of the system was much greater than the drift of the extremum caused by the variation of the characteristic curve $\eta=f(q)$ in time.

When a system contains no reactive components it is stable at all times. However, if the system contains inertial elements, its stability can be disturbed and it may fail to tune to the extremum.

Let us consider this problem in application to a step system, which measure at intervals of Δt seconds the extremal variable η and reverses when

$$\eta_i<\eta_{i-1}$$

To simplify the discussion, we shall assume that the reversal signal is produced when the instantaneous values of η, measured each Δt sec, are compared. In reality, the comparison is made between values averaged over the Δt intervals. We shall moreover assume that the extremum hunting device has no dead zone.

Therefore, unlike Figure 6.18 (the upper graph), the left-hand side of Figure 6.33 shows a continuous curve $\eta=f(q)$. We shall make a further assumption that the system in question has no linear output unit, i.e., no time lag arises at the output.

We shall start the analysis at the origin (Figure 6.33), where $q=0$. At the start, the input voltage u_1 (Figure 6.25) is $+k$, and q correspondingly increases. At points 1 and 2 the inequality $\eta_i<\eta_{i-1}$ is not satisfied. Finally at point 3 the inequality $\eta_i<\eta_{i-1}$ is satisfied and the voltage u_1 is consequently reversed. Since the system has a linear input unit introducing a time lag, the following two cases are possible.

a) Input time lag is very small

In this case (Figure 6.33) q will increase slightly beyond point 3 and then start rapidly decreasing. Following the decrease of q, the extremal variable η increases. Therefore, when the next measurement is taken (point 4), it is found that $\eta_i>\eta_{i-1}$, and the system will move on toward the extremum.

Hence, in spite of the time lag, the system will hunt normally about the extremum.

b) Input time lag is relatively large

If the time lag is large, q may persist in its upward trend owing to the inertia of the linear input unit (Figure 6.25), even after the reversal of u_1 at point 3.

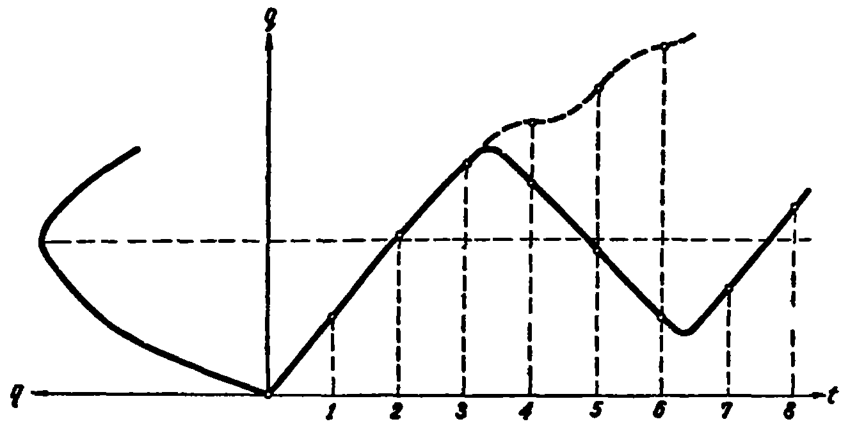

FIGURE 6.33. Characteristics of stable and unstable systems.

Therefore, when the next measurement is taken (point 4), it will be found that $\eta_i < \eta_{i-1}$ and the system will again be reversed, but this time in the wrong direction: q will increase instead of decreasing. The system will therefore recede from the extremum in an oscillatory manner (the dashed curve in Figure 6.33) and the stability of the system will be disturbed.

If there are time lags at both the input and the output of the system, its stability is best analyzed from the curve $u_2 = \varphi(q)$.

In a stable system (Figure 6.34), after reversal (point 3), q slightly increases and then starts decreasing (point 4^a). In an unstable system (dashed curve) q steadily increases in oscillatory motion beyond point 3.

An analysis of the curve $u_2 = \varphi(q)$ thus enables us to determine the stability of an adaptive control system and to analyze the dynamics of its tuning to the extremum.

13. HUNTING THE MINIMUM VALUE OF THE EXTREMAL VARIABLE

Hitherto only system where the controlled variable had a maximum, were considered. All the preceding can easily be applied to systems where

the minimum of the controlled variable is required. For example, Figure 6.35 shows the dynamic characteristic of a system hunting the minimum point of the extremal curve.

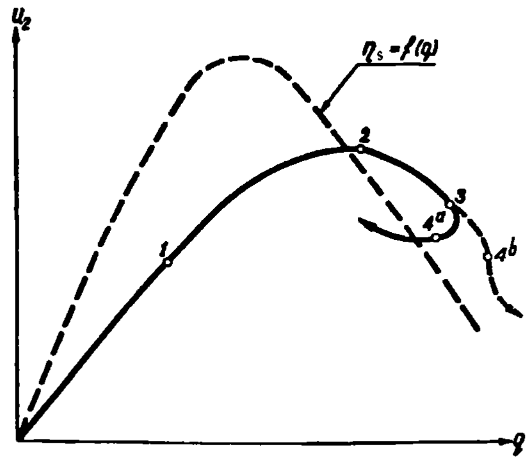

FIGURE 6.34. Dynamic characteristics of the system.

The function $\gamma = F(q)$ with a minimum (Figure 6.36) can easily be reduced to a curve $\eta = f(q)$ with a maximum. This transformation is attained by subtraction

$$\eta = u_k - \gamma, \tag{6.44}$$

where γ = the function with the minimum;
η = the function with the maximum;
u_k = constant voltage.

A regulator designed to hunt the maximum of a controlled variable (η) can thus be used to establish the minimum of a function (γ).

14. OPTIMIZATION OF FUNCTIONS WITH SEVERAL INDEPENDENT VARIABLES

In most common cases the extremal parameter (η) is a function of one variable (q) fed into the controlled system.

There are cases, however, when the extremal parameter is a function of one or more independent variables $(x, y, z, \ldots)$.

We shall consider some common extremization techniques for these cases.

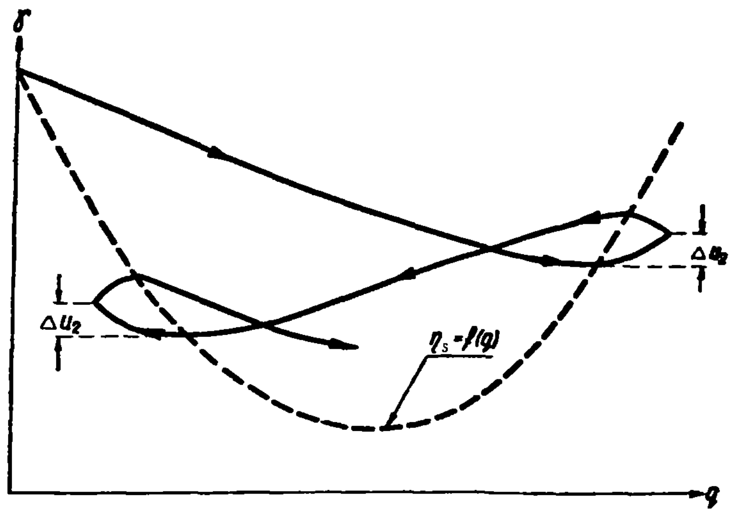

FIGURE 6.35. Dynamics of a system hunting its minimum.

a) Gauss-Seidel method
(successive variation of parameters)

This method calls for successive determination of partial extrema. In other words, an automatic-control system is first tuned to the extremum of the function

$$\eta = f_1(x).$$

When the system has adjusted itself to this extremum, it starts hunting for the second partial extremum:

$$\eta = f_2(y).$$

Following this, the functions

$$\eta = f_3(z)\ldots$$

are extremized.

The procedure is iterated until all the partial extrema are established. Then the system starts hunting again for the first extremum:

$$\eta = f_1(x).$$

The system thus operates in a cycle which involves the successive variation of the input parameters.

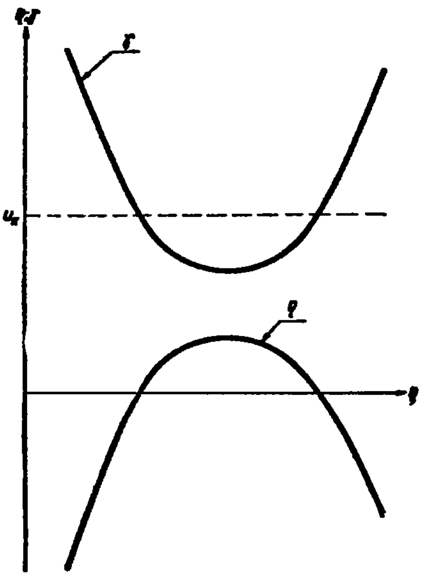

FIGURE 6.36. Transformation of a function with a minimum to a function with a maximum.

When the input of the system consists of n independent variables, the operation of the system is investigated in n-dimensional space. To simplify the problem, a system with only two independent variables (x and y) will be considered. In this case the process can be represented by plane curves.

The input variables x and y are laid off on the horizontal and vertical aixs (Figure 6.37). The origin is translated, as before (e.g., Figure 6.23), to the extremum.

In this plane we trace the loci of constant values for each extremal parameter; $\eta_1, \eta_2, \eta_3, \ldots, \eta_i \ldots$ Obviously,

$$\eta_i > \eta_3 > \eta_2 > \eta_1.$$

We further assume that at the start the system was located at point A. The system approaches the extremum (the origin) as follows.

First the variable x increases, so that for $y = y_1 =$ const, the function $\eta = f_1(x)$ is extremized. At this point (B), the system switches over to parameter y, which is subsequently varied. At point C, with $x = x_1 =$ const, the function $\eta = f_2(y)$ is extremized. The system again hunts for the extremum of the function $\eta = f_1(x)$, this time assuming a new value $y = y_2 =$ =const, closer to the origin than y_1, etc. The process is iterated until the system reaches the origin.

This method is convenient as it can be carried out with systems originally designed for extremizing functions of one variable. We see from Figure 6.37, however, that the system approaches the extremum by a path which is by no means the shortest.

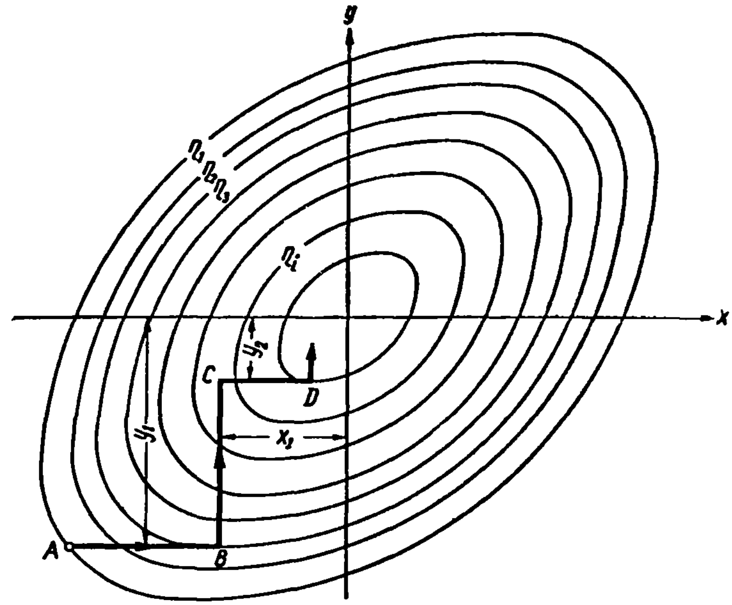

FIGURE 6.37. Dynamics of extremum tuning with successive variation of parameters.

b) The gradient method

This method of extremum tuning involves the simultaneous variation of all the independent parameters (x, y, z . .). The rate of change of each of these parameters (all of which change continuously) is proportional to the partial derivative of the extremal variable (η) with respect to the parameter concerned:

$$\frac{dx}{dt}=k\frac{\partial\eta}{\partial x};\ \frac{dy}{dt}=k\frac{\partial\eta}{\partial y};\ \frac{dz}{dt}=k\frac{\partial\eta}{\partial z}\cdots \tag{6.45}$$

If the parameters $x, y, z\ldots$ change discretely, the increment of each parameter is determined by the expressions

$$\Delta x=k\frac{\partial\eta}{\partial x};\ \Delta y=k\frac{\partial\eta}{\partial y};\ \Delta z=k\frac{\partial\eta}{\partial z}\cdots \tag{6.46}$$

The dynamics of extremum hunting for the two-variable case (x and y) is shown in Figure 6.38. In this figure it can be seen that the curve describing the motion of the system is at all times normal to the lines of equal values of the extremal variable (η).

When the gradient method is employed, the extremal parameter increases until the system reaches the extremum. The advantage of this method is the relatively short time required to maximize η, and the small hunting amplitude.

c) The brachistochrone method (method of fastest descent)

When situated at point A, the system establishes the normal to the surface $\eta = f(x, y, z \ldots)$ joining points with a constant value of η and passing through point A. The system then starts moving along this normal until the derivative of the function $\eta = f(x, y, z \ldots)$ in this direction vanishes. After that, the normal is again established and the system resumes its motion, etc.

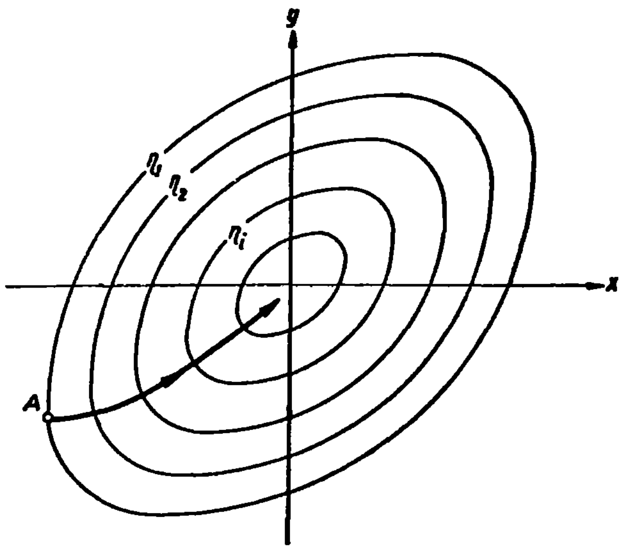

FIGURE 6.38. Dynamics of tuning to the extremum using the gradient method.

The path of the system in the case of two variables (x and y) is shown in Figure 6.39. The equation of the normal erected to the curve $f_1(x, y)$ at point A is

$$\frac{x - x_A}{\dfrac{\partial f_1(x,y)}{\partial x}} = \frac{y - y_A}{\dfrac{\partial f_1(x,y)}{\partial y}}, \tag{6.47}$$

where x_A, y_A = the coordinates of point A;

$f_1(x, y)$ = the loci of constant values of η passing through point A.

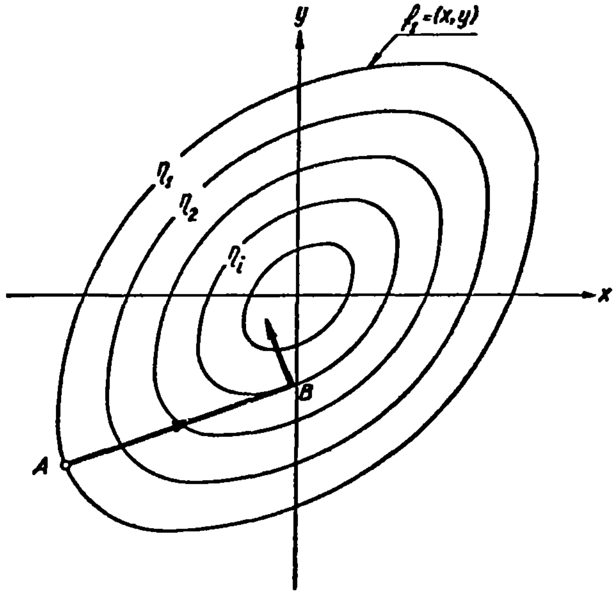

FIGURE 6.39. Dynamics of extremum tuning with the brachistochrone method.

When the normal direction has been fixed, the system starts moving along this normal. At point *B*, where the extremal parameter (η) is maximized, the system establishes the normal to the loci of constant of passing through this new point. This procedure is repeated until the system reaches the origin.

The advantage of the brachistochrone method is the relatively short time required to approach the extremum.

Chapter VII

LEARNING IN AUTOMATIC–CONTROL SYSTEMS

1. THE PROBLEM OF LEARNING

The application of learning in automatic-control systems is the climax in the development of engineering cybernetics.

Learning systems, unlike adaptive systems, are at first incapable of performing any control functions whatsoever. They collect information on the controlled object and its behavior. They as if observe the process of manual control, or control by some other automatic device. Only after these systems "learn" how to control, can they take over. The process of learning does not stop here. While carrying out active control, the systems still "investigate" the controlled object and modify the control specifications so that the control characteristics are improved.

The ability of living organisms to maintain certain characteristics in spite of considerable variation of external conditions is called homeostasis*. Ashby designed an apparatus demonstrating how learning introduced changes in the reaction of the apparatus to varying external conditions. This apparatus was called a homeostat.

Ashby's homeostat consisted of four electromagnets suspended on hinges. A container with electrolyte was placed under each magnet. Electrodes were provided along the edge of each container, and a rod immersed in the liquid was attached to each magnet. Each magnet, together with the container, thus comprised a potentiometer which varied the voltage depending on the angle of inclination of the hinged electromagnet.

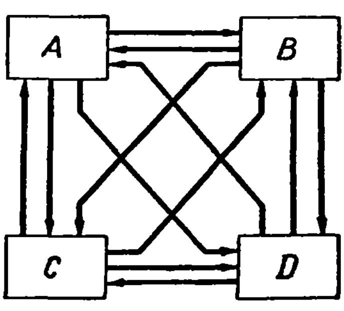

FIGURE 7.1. Block diagram of Ashby's homeostat.

Amplifiers were connected to each potentiometer giving four identical units (Figure 7.1). These units were so connected that the output voltage of each was fed to the other three units. The input of each unit thus received voltage from all the other units. In addition a feedback loop fed the output voltage of each unit to its own input.

The torque in each unit was thus proportional to the output voltage of all four units.

Switches were provided at the inputs of each unit to change (discretely) the polarity and gain of each voltage. There was a total of 390,625 different combinations of the

* ὁμοιόστασις — the state of stable equilibrium.

parameters. The number of variables was eight (the angle of inclination of the four magnets and the position of the four switches).

The object of this apparatus, according to Ashby, was to produce "ultrastability". A stable state existed when all the four suspended magnets were in the vertical (zero) position. This state had to be maintained under very different external conditions: coils open or short-circuited, polarity of the magnets reversed, the resistances introduced into the coil circuits, etc.

The homeostat was so built that, when it was disturbed by an external stimulus, e.g., a short-circuit in one of the coils, the selector switches stepped over all the 390,625 positions and selected those positions where the voltages of all four electromagnets were zero.

Strictly speaking, the term "learning" does not apply to Ashby's homeostat, since this device only selected coil polarities and gain factors to ensure stability under new, modified conditions. The homeostat returned to the initial state (when the initial conditions were restored) only after random searching (processing hundreds of thousands of possible combinations).

Shannon's so-called "mouse" is very famous. This "mouse" moved in a maze of previously unknown layout and felt the walls of the maze as it hit them. After many a trial and error, it emerged at the correct end of the maze. In its first run through the maze, the "mouse" as if "learned" the layout. Therefore, when the "mouse" ran again through the maze, it traced the correct path without hitting the walls.

It must be kept in mind that the process of remembering and forgetting in living organisms and in computing devices follows essentially different patterns. Computing devices "memorize" information almost instantaneously and "store" it until a new command makes them forget. Animals, on the other hand, learn slowly and forget slowly.

A learning system must therefore be designed on other principles than those used in computor engineering.

Quite interesting in this respect was the device developed by Walter and called by its creator "Cora". "Cora" was built as a small trolley which "hid" (rolled aside) when pushed, but at first did not react to sound. If the push was accompanied several times by sound, "Cora" started reacting (rolled aside) also to this new signal. However, if "Cora" was pushed several times without the accompaniment of sound, it "forgot" the meaning of sound and again reacted only to the mechanical stimulus. "Cora" thus approximately imitated the development of a conditioned reflex. We shall now consider this problem in more detail.

2. IMITATION OF THE LEARNING PROCESS

Imitating the process of development of conditioned reflexes in automatic-control systems is of prime importance for the design of learning systems.

Pavlov showed that if a bell was sounded a sufficient number of times before an animal received its food, then, after a certain period, the animal reacted to this sound as if it were food, i.e., when the bell rang, saliva was secreted.

Consider simulating this experiment by means of the circuit shown in Figure 7.2.

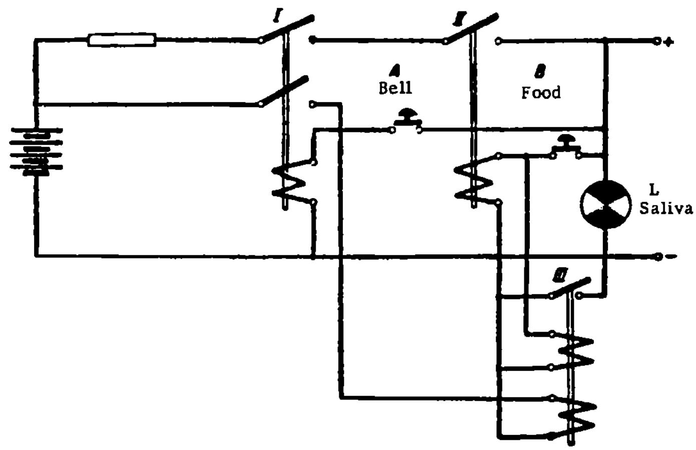

FIGURE 7.2. Electrical simulation of a conditioned reflex.

In this circuit when button A is pressed ("the bell rings"), the "saliva" signal does not light, since in this case only contact I is made (the battery is not charged). If now button B is pressed ("give food"), switch III is closed and the signal lamp lights.

Let us now press buttons A and B simultaneously (the bell rings when the food is given). In this case, contacts I and II are made simultaneously and the battery receives some charge. After repeated pressing of the two buttons the battery is sufficiently charged, so that when only button A (the bell) is pressed, contact III also makes, and the "saliva" signal lights.

However, this situation will not persist indefinitely. If now button A is pressed several times in succession, the battery discharges. After that the system will not react to signal A.

The circuit shown in Figure 7.2 simulates the development of a conditioned reflex and is thus considered a learning element. A learning system based on this element measures the probability of coincidence of events A and B. If this probability is high, the battery is gradually charged. If the probability is low, the battery does not charge (or even discharge).

Let us consider still another system (Figure 7.3) assembled of basic logic elements.

Voltage is periodically applied to inputs (A and B) of the system (by analogy with the pressing of buttons A and B in Figure 7.2). The AND_1 gate enables counter C_1 to count the number of simultaneous occurrences of signals A and B (the analogy in Figure 7.2 is the charging of the battery when the two buttons are pressed simultaneously). Counter C_2 records the number of times signal A is not accompanied by signal B (the battery-discharging analogy in Figure 7.2).

Counter C_1 emits an output signal after n_1 counts. Similarly, counter C_2 produces an output signal after receiving n_2 counts. Each counter is reset to zero when an input signal is received by the other counter. The counter outputs are fed into a trigger circuit made of four gates (OR_1, NOT_2, OR_2, NOT_3). We see from the block diagram that C_1 sets point a to 1, and counter C_2 clears it to zero.

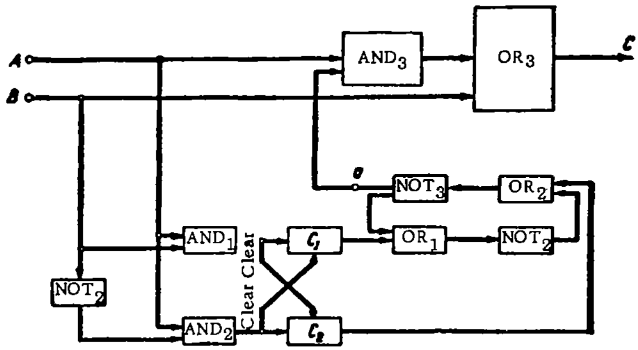

FIGURE 7.3. Logic element system simulating a conditioned reflex.

The system operates as follows. When signals A and B coincide, the AND_1 gate produces an output signal, the counter C_1 keeping record of these coincidence signals. When the accumulated number is n_1, the counter feeds a signal into the trigger circuit and point a is set to 1.

At first, if the system receives signal A only, the output of the system is 0. However, after point a is set to 1, signal A alone (like signals A and B together) produces an output of 1 at C.

If signal A is not accompanied by signal B, the gates NOT_1, AND_2 are actuated, and clear to zero counter C_1. Simultaneously, counter C_2 starts counting the number of separate occurrences of signal A, when not accompanied by signal B. If this number is equal to n_2, counter C_2 emits an output signal, the trigger circuit is actuated, and point a is reset to 0. After this signal, A alone will not produce an output signal at C.

If, as C_2 counts, an input signal is received by C_1, counter C_2 is cleared. The system in Figure 7.3 thus operates analogously to the circuit in Figure 7.2.

Note that the system in Figure 7.3, as in the previous circuit, changes its logical operation depending on the probability of simultaneous occurrence of signals A and B. Indeed, it follows from Figure 7.3 that when the probability of coincidence of signals A and B is low, the system performs the following logical proposition:

$$C=B.$$

If the coincidence probability is high, the system performs a different logical proposition:

$$C=A+B.$$

The circuits in Figures 7.2 and 7.3 are in fact the basic elements used in composite learning automatic-control systems.

Let us consider a simple example in the application of the system in Figure 7.3. The cutting depth of a cylindrical part by a lathe is to be automatically regulated. The cutter in this lathe can be adjusted so that the part in question is turned to the required diameter. However, as the

cutter wears out, the final diameter of the machined parts gradually increases, until eventually rejects appear whose diameters are greater than the permissible maximum. To avoid this, the cutter must be moved toward the center of the machined part so as to compensate for its wear.

On the other hand, the cutter meets different resistances from different parts because the parts have not been cast simultaneously, or have acquired individual corrosion crusts, etc. In addition, the backlash and play of the cutter holder, cause the diameter of the finished product to be a random variable. The increase of the diameter of one (k-th) part when compared with the diameter of the preceding (k-1-th) part cannot be interpreted as a signal for cutter adjustment, since the diameter of the next (k+1-th) part can again be smaller than the diameter of the present part. Hence the cutter must be adjusted when the probability that the diameter has increased, because of cutter wear, is sufficiently high.

Applying the principle of the system in Figure 7.3, we set up the circuit in Figure 7.4 for the automatic control of the lathe cutter.

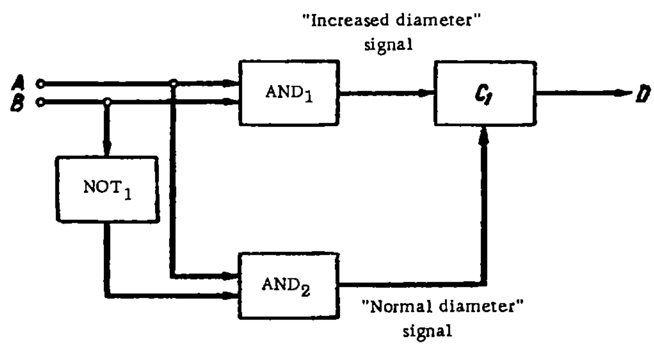

FIGURE 7.4. Logical circuit for selector control.

Two signals are received at the input of this system: signal A indicating a part is being machined, and signal B indicating that the diameter of the machined part is greater than a preassigned value. This control value is taken smaller than the permissible maximum, so that a random increase in diameter will not cause rejection.

The AND_1 gate thus produces an output signal whenever the diameter of the machined part is greater than the preassigned value. Counter C_1 keeps record of the number of these parts and emits a signal when this number, as in Figure 7.3, reaches n_1.

The AND_2 gate produces an output signal when the diameter of the machined part is not greater than the preassigned value. This signal clears counter C_1 to zero. A signal at the counter output thus indicates that the probability of occurrence of parts with increased diameter has become sufficiently high. This signal is used by a selector to rotate the cutter guide screw, adjusting the cutter to the next prescribed position.

In this problem, it is naturally assumed that the distance over which the cutter is displaced, when the selector moves one position, is sufficient

to make the diameter of the machined parts smaller than the prescribed value.

Let us consider an example when the investigated signal is continuous.

Take two batteries (Figure 7.5), one of which is being charged and the other discharged. Each battery is switched from one mode to the other by means of a selector. The battery being discharged must be switched over to charge when its voltage drops to a prescribed level, and the other battery, which until then has been charging, must be switched over to discharge.

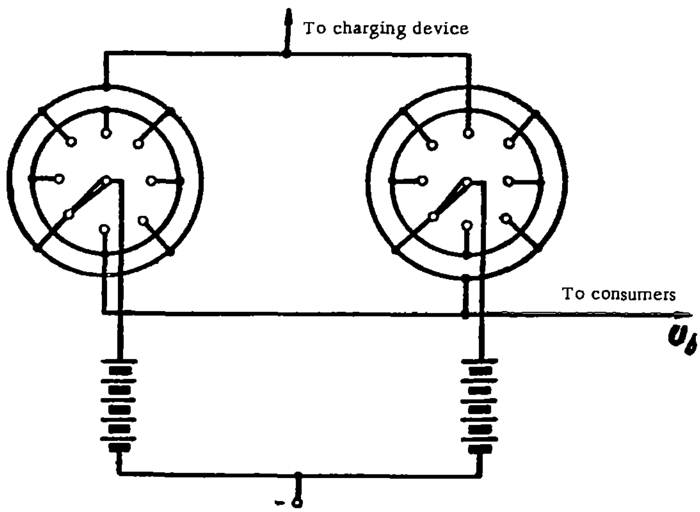

FIGURE 7.5. Two-battery control circuit.

It should be observed that the batteries supply various consumers, whose characteristics and operating curves are generally unknown. Moreover, when an electrical motor is started, the battery voltage briefly drops. The battery voltage (Figure 7.6) is therefore a random function of time.

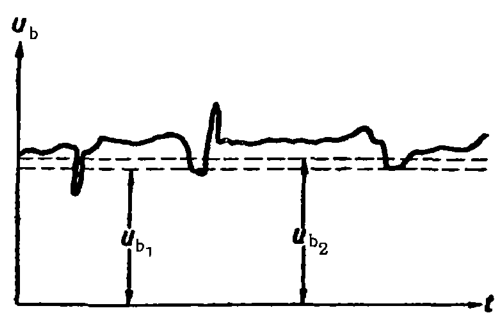

FIGURE 7.6. Battery voltage.

Hence in this example the batteries should not be switched over immediately when the voltage of the discharging battery reaches some minimum value (as is usually done in ordinary error-regulating systems).

Using the principle of Figure 7.2, the circuit in Figure 7.7 was designed for the switching of batteries.

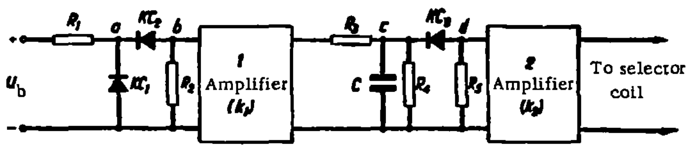

FIGURE 7.7. Control circuit for battery switching.

The operation of the individual elements of this system is shown in Figure 7.8. The first element (Figure 7.8,a) comprising a resistor R_1 and a Zener diode KC_1 clips those parts of the continuous voltage curve u_b where the voltage is greater than some prescribed value u_{b_2}.

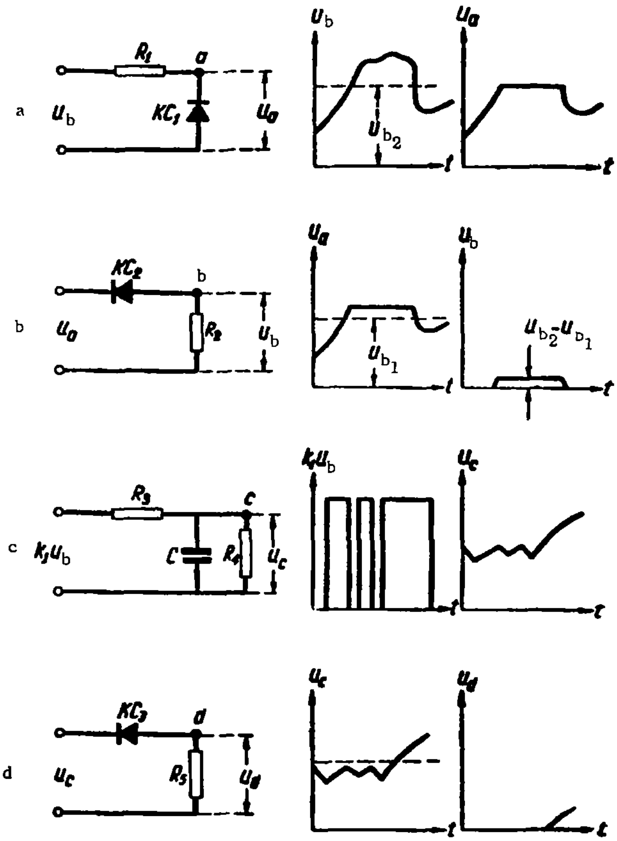

FIGURE 7.8. Logical operation.

The second element (Figure 7.8,b) consisting of a Zener diode KC_2 and a resistor R_2 as if lowers the maximum voltage u_a (that which exceeds u_{b_1}) to the abscissa axis (voltage u_b)* .

The input of the first amplifier thus consists of almost rectangular impulses. The ratio of the duration of these impulses to the overall time of measurement gives the probability that the battery voltage falls in the range $u_{b_2} > u_b > u_{b_1}$.

To control the selector in Figure 7.5, this probability need not be computed. It suffices to determine the point at which the probability P attains a prescribed value P_1. We therefore proceed as follows.

Voltage u_b is amplified and applied to an element (Figure 7.8,c) consisting of a capacitor (C) and two resistors (R_3 and R_4). When an output voltage is produced by this element, the capacitor charges, and when there is no output voltage, it discharges (through resistor R_4). The probability P_1 can therefore be characterized by the output voltage u_c of this element.

Hence, clipping the lower part of the curve $u_c = f(t)$ by means of the element shown in Figure 7.8,c, we obtain the voltage u_d representing the probability that the voltage of the battery falls in the range $u_{b_2} > u_b > u_b$ is greater than the prescribed value P_1. If now voltage u_d is amplified and fed to the selector coil, the battery is automatically switched over to the charging regime, as required.

At present we are witnessing the first stages in the application of learning systems to various branches of the national economy. The existing automatic-control systems are capable of imitating only the simplest elements of the learning process in animals. However, the vigorous development of cybernetics shows that this imitation is continually improving from year to year. There is no doubt that in the near future learning systems will find a very wide range of applications.

* *
*

Because of the comparatively small scope of this book we naturally could not cover the entire range of problems and topics advanced by engineering cybernetics.

This, however, was not the aim of the book. If the reader, having worked his way through the book, has acquired some interest in engineering cybernetics and learned how cybernetics can be applied to the solution of practical problems, then the author's object has been achieved. For further details the reader should refer to specialized scientific literature.

* [This element acts as a discriminator.]

BIBLIOGRAPHY

1. Automata. — Collection of articles, edited by C. Shannon and D. MacCarty. Russian translation. 1956.
2. Berckley, E. Symbolic Logic and Intelligent Machines. Russian translation. 1961.
3 Ventsel', E.S. Teoriya veroyatnostei (Theory of Probability). — Fizmatgiz. 1962.
4. Wiener, N. Cybernetics. Russian translation. 1958.
5. Gaaze—Rapoport, M.G. Avtomaty i zhivye organizmy (Automata and Living Organisms). — Fizmatgiz. 1961.
6. Ivakhnenko, A.G. Tekhnicheskaya kibernetika (Engineering Cybernetics). — Izdatel'stvo Tekhnicheskoi Literatury Ukr. SSR. 1959.
7. Klein, M., H. Morgan, and M. Ahronson. Digital Technique for Computations and Control. Russian translation. 1960.
8. Cossat, P. Cybernetics. Russian translation. 1958.
9. Osnovy avtomaticheskogo regulirovaniya (Principles of Automatic Control). Ed. by V.V. Solodovnikov. — Mashgiz. 1954.
10. Solodovnikov, V.V. Vvedenie v statisticheskuyu dinamiku sistem avtomaticheskogo regulirovaniya (Introduction to Statistical Dynamics of Automatic-control Systems). —Gosenergoizdat. 1952.
11. Truxall, D. Design of Automatic-control Systems. Russian translation. 1959.
12. Fel'dbaum, A.A. Elektricheskie sistemy avtomaticheskogo regulirovaniya (Electrical Automatic-control Systems). — Oborongiz. 1957.
13. Fel'dbaum, A.A. Vychislitel'nye ustroistva v avtomaticheskikh sistemakh (Computing Devices in Automatic Systems). — Fizmatgiz. 1959.
14.. T'sin Hu-Sen. Engineering Cybernetics. Russian translation. 1956.
15. Ashby, W. Introduction to Cybernetics. Russian translation. 1959.

SUBJECT INDEX